Licheng Jiao
Lipo Wang
Xinbo Gao
Jing Liu
Feng Wu (Eds.)

LNCS 4222

Advances in Natural Computation

Second International Conference, ICNC 2006
Xi'an, China, September 2006
Proceedings, Part II

2 Part II

Springer

Lecture Notes in Computer Science 4222

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Licheng Jiao Lipo Wang Xinbo Gao
Jing Liu Feng Wu (Eds.)

Advances in Natural Computation

Second International Conference, ICNC 2006
Xi'an, China, September 24-28, 2006
Proceedings, Part II

Springer

Volume Editors

Licheng Jiao
Xidian University, Xi'an 710071, China
E-mail: lchjiao@mail.xidian.edu.cn

Lipo Wang
Nanyang Technological University, Singapore
E-mail: elpwang@ntu.edu.sg

Xinbo Gao
Xidian University, Xi'an, 710071 China
E-mail: xbgao@mail.xidian.edu.cn

Jing Liu
Xidian University, Xi'an, 710071, China
E-mail: neouma@mail.xidian.edu.cn

Feng Wu
Microsoft Research Asia, Haidian, 100080 Beijing, China
E-mail: fengwu@microsoft.com

Library of Congress Control Number: 2006933052

CR Subject Classification (1998): F.1, F.2, I.2, G.2, I.4, I.5, J.3, J.4

LNCS Sublibrary: SL 1 – Theoretical Computer Science and General Issues

ISSN 0302-9743
ISBN-10 3-540-45907-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-45907-1 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11881223 06/3142 5 4 3 2 1 0

Preface

This book and its sister volumes, i.e., LNCS volumes 4221 and 4222, constitute the proceedings of the 2nd International Conference on Natural Computation (ICNC 2006), jointly held with the 3rd International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2006, LNAI volume 4223) on 24-28 September 2006 in Xi'an, Shaanxi, China. In its budding run, ICNC 2006 successfully attracted 1915 submissions from 35 countries/regions (the joint ICNC-FSKD 2006 event received 3189 submissions). After rigorous reviews, 254 high-quality papers, i.e., 168 long papers and 86 short papers, were included in the ICNC 2006 proceedings, representing an acceptance rate of 13.3%.

ICNC-FSKD 2006 featured the most up-to-date research results in computational algorithms inspired from nature, including biological, ecological, and physical systems. It is an exciting and emerging interdisciplinary area in which a wide range of techniques and methods are being studied for dealing with large, complex, and dynamic problems. The joint conferences also promoted cross-fertilization over these exciting and yet closely-related areas, which had a significant impact on the advancement of these important technologies. Specific areas included neural computation, quantum computation, evolutionary computation, DNA computation, fuzzy computation, granular computation, artificial life, etc., with innovative applications to knowledge discovery, finance, operations research, and more. In addition to the large number of submitted papers, we were blessed with the presence of six renowned keynote speakers.

On behalf of the Organizing Committee, we thank Xidian University for sponsorship, and the National Natural Science Foundation of China, the International Neural Network Society, the Asia-Pacific Neural Network Assembly, the IEEE Circuits and Systems Society, the IEEE Computational Intelligence Society, the IEEE Computational Intelligence Singapore Chapter, and the Chinese Association for Artificial Intelligence for technical co-sponsorship. We thank the members of the Organizing Committee, the Advisory Board, and the Program Committee for their hard work in the past 12 months. We wish to express our heartfelt appreciation to the keynote speakers, session chairs, reviewers, and student helpers. Our special thanks go to the publisher, Springer, for publishing the ICNC 2006 proceedings as two volumes of the Lecture Notes in Computer Science series (and the FSKD 2006 proceedings as one volume of the Lecture Notes in Artificial Intelligence series). Finally, we thank all the authors and participants for their great contributions that made this conference possible and all the hard work worthwhile.

September 2006

Lipo Wang
Licheng Jiao

Organization

ICNC 2006 was organized by Xidian University and technically co-sponsored by the National Natural Science Foundation of China, the International Neural Network Society, the Asia-Pacific Neural Network Assembly, the IEEE Circuits and Systems Society, the IEEE Computational Intelligence Society, the IEEE Computational Intelligence Singapore Chapter, and the Chinese Association for Artificial Intelligence.

Organizing Committee

Honorary Conference Chairs:	Shun-ichi Amari (RIKEN BSI, Japan)
	Xin Yao (University of Birmingham, UK)
General Co-chairs:	Lipo Wang (Nanyang Technological University, Singapore)
	Licheng Jiao (Xidian University, China)
Program Committee Chairs:	Xinbo Gao (Xidian University, China)
	Feng Wu (Microsoft Research Asia, China)
Local Arrangement Chairs:	Yuanyuan Zuo (Xidian University, China)
	Xiaowei Shi (Xidian University, China)
Proceedings Chair:	Jing Liu (Xidian University, China)
Publicity Chair:	Yuping Wang (Xidian University, China)
Sponsorship Chair:	Yongchang Jiao (Xidian University, China)
Secretaries:	Bin Lu (Xidian University, China)
	Tiantian Su (Xidian University, China)
Webmasters:	Yinfeng Li (Xidian University, China)
	Maoguo Gong (Xidian University, China)

Advisory Board

Zheng Bao	Xidian University, China
Zixing Cai	Central South University, China
Guoliang Chen	University of Science and Technology of China, China
Huowang Chen	National University of Defense Technology, China
David Corne	The University of Exeter, UK
Dipankar Dasgupta	University of Memphis, USA
Kalyanmoy Deb	Indian Institute of Technology Kanpur, India
Baoyan Duan	Xidian University, China
Kunihiko Fukushima	Tokyo University of Technology, Japan
Tom Gedeon	The Australian National University, Australia

Aike Guo	Chinese Academy of Science, China
Yao Hao	Xidian University, China
Zhenya He	Southeastern University, China
Fan Jin	Southwest Jiaotong University, China
Yaochu Jin	Honda Research Institute Europe, Germany
Janusz Kacprzyk	Polish Academy of Sciences, Poland
Lishan Kang	China University of Geosciences, China
Nikola Kasabov	Auckland University of Technology, New Zealand
John A. Keane	The University of Manchester, UK
Soo-Young Lee	KAIST, Korea
Yanda Li	Tsinghua University, China
Zhiyong Liu	National Natural Science Foundation of China, China
Erkki Oja	Helsinki University of Technology, Finland
Nikhil R. Pal	Indian Statistical Institute, India
Yunhe Pan	Zhe Jiang University, China
Jose Principe	University of Florida, USA
Witold Pedrycz	University of Alberta, Canada
Marc Schoenauer	University of Paris Sud, France
Zhongzhi Shi	Chinese Academy of Science, China
Harold Szu	Office of Naval Research, USA
Shiro Usui	RIKEN BSI, Japan
Shoujue Wang	Chinese Academy of Science, China
Xindong Wu	University of Vermont, USA
Lei Xu	Chinese University of Hong Kong, HK
Bo Zhang	Tsinghua University, China
Nanning Zheng	Xi'an Jiaotong University, China
Yixin Zhong	University of Posts & Telecommunications, China
Syozo Yasui	Kyushu Institute of Technology, Japan
Jacek M. Zurada	University of Louisville, USA

Program Committee

Shigeo Abe	Kobe University, Japan
Davide Anguita	University of Trento, Italy
Abdesselam Bouzerdoum	University of Wollongong, Australia
Laiwan Chan	The Chinese University of Hong Kong, HK
Li Chen	Northwest University, China
Guanrong Chen	City University of Hong Kong, HK
Shu-Heng Chen	National Chengchi University, Taiwan
Tianping Chen	Fudan University, China
YanQiu Chen	Fudan University, China
Vladimir Cherkassky	University of Minnesota, USA

Sung-Bae Cho	Yonsei University, Korea
Sungzoon Cho	Seoul National University, Korea
Tommy W.S. Chow	City University of Hong Kong, China
Vic Ciesielski	RMIT, Australia
Keshav Dahal	University of Bradford, UK
L.N. de Castro	Catholic University of Santos, Brazil
Emilio Del-Moral-Hernandez	University of Sao Paulo, Brazil
Andries Engelbrecht	University of Pretoria, South Africa
Tomoki Fukai	Tamagawa University, Japan
Lance Fung	Murdoch University, Australia
Takeshi Furuhashi	Nagoya University, Japan
Hiroshi Furutani	University of Miyazaki, Japan
John Q. Gan	The University of Essex, UK
Wen Gao	The Chinese Academy of Science, China
Peter Geczy	AIST, Japan
Zengguang Hou	University of Saskatchewan, Canada
Jiwu Huang	Sun Yat-Sen University, China
Masumi Ishikawa	Kyushu Institute of Technology, Japan
Yongchang Jiao	Xidian University, China
Robert John	De Montfort University, UK
Mohamed Kamel	University of Waterloo, Canada
Yoshiki Kashimori	University of Electro-Communications, Japan
Samuel Kaski	Helsinki University of Technology, Finland
Andy Keane	University of Southampton, UK
Graham Kendall	The University of Nottingham, UK
Jong-Hwan Kim	KAIST, Korea
JungWon Kim	University College London, UK
Natalio Krasnogor	University of Nottingham, UK
Vincent C.S. Lee	Monash University, Australia
Stan Z. Li	Chinese Academy of Science, China
Yangmin Li	University of Macau, Macau
Xiaofeng Liao	Chongqing University, China
Derong Liu	University of Illinois at Chicago, USA
Ding Liu	Xi'an University of Technology, China
Jing Liu	Xidian University, China
Ke Liu	National Natural Science Foundation of China, China
Baoliang Lu	Shanghai Jiao Tong University, China
Frederic Maire	Queensland University of Technology, Australia
Jacek Mandziuk	Warsaw University of Technology, Poland
Satoshi Matsuda	Nihon University, Japan
Masakazu Matsugu	Canon Research Center, Japan
Bob McKay	University of New South Wales, Australia
Ali A. Minai	University of Cincinnati, USA
Hiromi Miyajima	Kagoshima University, Japan
Hongwei Mo Harbin	Engineering University, China

Mark Neal	University of Wales, Aberystwyth, UK
Pedja Neskovic	Brown University, USA
Richard Neville	The University of Manchester, UK
Tohru Nitta	National Institute of Advanced Industrial Science and Technology, Japan
Yusuke Nojima	Osaka Prefecture University, Japan
Takashi Omori	Hokkaido University, Japan
Yew Soon Ong	Nanyang Technological University, Singapore
M. Palaniswami	The University of Melbourne, Australia
Andrew P. Paplinski	Monash University, Australia
Asim Roy	University of Arizona, USA
Bernhard Sendhoff	Honda Research Centre Europe, Germany
Leslie Smith	University of Stirling, UK
Andy Song	RMIT, Australia
Lambert Spaanenburg	Lund University, Sweden
Changyin Sun	Southeast University, China
Mingui Sun	University of Pittsburgh, USA
Johan Suykens	KULeuven, Belgium
Kay Chen Tan	National University of Singapore, Singapore
Jonathan Timmis	University of York, UK
Seow Kiam Tian	Nanyang Technological University, Singapore
Peter Tino	The University of Birmingham, UK
Kar-Ann Toh	Institute of Infocomm Research, Singapore
Yasuhiro Tsujimura	Nippon Institute of Technology, Japan
Ganesh Kumar Venayagamoorthy	University of Missouri-Rolla, USA
Ray Walshe	Dublin City University, Ireland
Lei Wang	Xi'an University of Technology, China
Xiaofan Wang	Shanghai Jiaotong University, China
Xufa Wang	University of Science and Technology of China, China
Yuping Wang	Xidian University, China
Sumio Watanabe	Tokyo Institute of Technology, Japan
Gang Wei	South China University of Technology, China
Stefan Wermter	University of Sunderland, UK
Kok Wai Wong	Murdoch University, Australia
Feng Wu	Microsoft Research Asia, China
Xihong Wu	Peking University, China
Zongben Xu	Xi'an Jiaotong University, China
Ron Yang	University of Exeter, UK
Li Yao	Beijing Normal University, China
Daniel Yeung	The Hong Kong Polytechnic University, HK
Ali M.S. Zalzala	Heriot-Watt University, UK
Hongbin Zha	Peking University, China
Liming Zhang	Fudan University, China
Qingfu Zhang	The University of Essex, UK

Wenxiu Zhang	Xi'an Jiaotong University, China
Yanning Zhang	Northwestern Polytechnical University, China
Yi Zhang	University of Electronic Science and Technology of China, China
Zhaotian Zhang	National Natural Science Foundation of China, China
Liang Zhao	University of Sao Paulo, Brazil
Mingsheng Zhao	Tsinghua University, China
Qiangfu Zhao	University of Aizu, Japan

Reviewers

A. Attila ISLIER
Abdesselam BOUZERDOUM
Adel AZAR
Ah-Kat TAN
Aifeng REN
Aifeng REN
Ailun LIU
Aimin HOU
Aizhen LIU
Ales KEPRT
Ali Bekir YILDIZ
Alparslan TURANBOY
Anan FANG
Andreas HERZOG
Andrew TEOH
Andrew P. PAPLINSKI
Andries ENGELBRECHT
Andy SONG
Anni CAI
Ariel GOMEZ
Arumugam S.
Ay-Hwa A. LIOU
Bang-Hua YANG
Baolong GUO
Bei-Ping HOU
Bekir CAKIR
Ben-Shun YI
Bharat BHASKER
Bin XU
Bin JIAO
Bin LI
Bing HAN
Binghai ZHOU
Bo FU
Bob MCKAY
Bohdan MACUKOW
Bo-Qin FENG
Brijesh VERMA
Caihong MU
Ce FAN
Changyin SUN
Changzhen HU
Chao DENG
Chaojian SHI
Chaowan YIN
Chen YONG
Cheng WANG
Chengxian XU
Cheng-Yuan CHANG
Cheol-Hong MOON
Chi XIE
Ching-Hung LEE
Chong FU
Chonghui GUO
Chong-Zhao HAN
Chor Min TAN
Chu WU
Chuang GUO
Chuanhan LIU
Chun JIN
Chun CHEN
Chung-Li TSENG
Chunshien LI
Cong-Kha PHAM
Cuiqin HOU
Cunchen GAO

Daehyeon CHO
Dat TRAN
Davide ANGUITA
De XU
Deqin YAN
Dewu WANG
Dexi ZHANG
Deyun CHEN
Diangang WANG
Dong LIU
Dong Hwa KIM
Dongbo ZHANG
Dongfeng HAN
Donghu NIE
Dong-Min WOO
Du-Yun BI
Emilio DEL-MORAL-HERNANDEZ
En-Min FENG
Ergun ERASLAN
Euntai KIM
Fajun ZHANG
Fang LIU
Fangshi WANG
Fan-Hua YU
Fei HAO
Fei GAO
Feng SHI
Feng XUE
Feng DING
Feng CHEN
Feng JIAO
Feng GAO
Fenlin LIU
Fu-Ming LI
Gabriel CIOBANU
Gang WANG
Gang CHEN
Gaofeng WANG
Gaoping WANG
Gary YEN
Gexiang ZHANG
Golayoglu Fatullayev AFET
Graham KENDALL
Guang REN
Guang TIAN
Guang LI
Guangming SHI
Guangqiang LI
Guang-Qiu HUANG
Guangrui WEN
Guang-Zhao CUI
Guanjun WANG
Guanlong CHEN
Guanzheng TAN
Gui-Cheng WANG
Guixi LIU
Guojun ZHANG
Guowei YANG
Guoyin WANG
Guo-Zheng LI
Gurvinder BAICHER
Gwi-Tae PARK
Hai TAO
Hai-Bin DUAN
Haifeng DU
Haiqi ZHENG
Haixian WANG
Haixiang GUO
Haiyan JIN
Hajime NOBUHARA
Hanjun JIN
Hao WANG
Haoran ZHANG
Haoyong CHEN
He JIANG
Hengqing TONG
Hiroshi FURUTANI
Hong JIN
Hong ZHANG
Hong LIU
Hong Jie YU
Hongan WANG
Hongbin DONG
Hongbing JI
Hongcai ZHANG
Honghua SHI
Hongsheng SU
Hongwei MO
Hongwei LI
Hongwei SI

Hongwei HUO
Hongxin ZHANG
Hongyu LI
Hongzhang JIN
Hua-An ZHAO
Huaxiang LU
Hua-Xiang WANG
Huayong LIU
Hui YIN
Hui LI
Hui WANG
Huizhong YANG
Hyun YOE
Hyun Chan CHO
Hyun-Cheol JEONG
Ihn-Han BAE
Ilhong SUH
In-Chan CHOI
I-Shyan HWANG
Ivan Nunes Da SILVA
Jae Hung YOO
Jae Yong SEO
Jae-Jeong HWANG
Jae-Wan LEE
Jea Soo KIM
Jia LIU
Jiafan ZHANG
Jian YU
Jian SHI
Jian CHENG
Jian XIAO
Jianbin SONG
Jiang CUI
Jiangang LU
Jianguo JIANG
Jianhua PENG
Jianjun WANG
Jianling WANG
Jian-Sheng QIAN
Jianwei YIN
Jianwu DANG
Jianyuan JIA
Jiating LUO
Jidong SUO
Jie LI
Jie HU
Jie WANG
Jie LI
Jih-Chang HSIEH
Jih-Fu TU
Jih-Gau JUANG
Jili TAO
Jil-Lin LIU
Jin YANG
Jinchao LI
Jinfeng YANG
Jing LIU
Jing-Min WANG
Jingwei LIU
Jingxin DONG
Jin-Ho KIM
Jinhui ZHANG
Jinling ZHANG
Jinping LI
Jintang YANG
Jin-Young KIM
Jiqing QIU
Jiquan SHEN
Ji-Song KOU
Jiu-Chao FENG
Jiulong ZHANG
Jiuying DENG
Jiyang DONG
Jiyi WANG
Johan SUYKENS
John Q GAN
Jong-Min KIM
Joong-Hwan BAEK
Jorge CASILLAS
Jose SEIXAS
Jr-Syu YANG
Ju Cheng YANG
Ju Han KIM
Juan LIU
Jumin ZHAO
Jun GAO
Jun YANG
Jun CAO
Jun JING
Jun MENG

Jun-An LU
Jung-Hyun YUN
Jungsik LEE
Junguo SUN
Junping ZHANG
Jun-Seok LIM
Jun-Wei LU
Junyi SHEN
Junying ZHANG
Kay Chen TAN
Kay-Soon LOW
Ke LU
Kefeng FAN
Kenneth REVETT
Keun-Sang PARK
Khamron SUNAT
Kok Wai WONG
Kwan Houng LEE
Kwang-Baek KIM
Kyung-Woo KANG
Laicheng CAO
Laiwan CHAN
Lambert SPAANENBURG
Lan GAO
Lance FUNG
Lean YU
Lei LIN
Lei WANG
Leichun WANG
Li MEIJUAN
Li WU
Li DAYONG
Li SUN
Li ZHANG
Liang GAO
Liang XIAO
Liang MING
Lian-Wei ZHAO
Lianxi WU
Liefeng BO
Lili ZHOU
Liming CHEN
Li-Ming WANG
Lin CONG
Lincheng SHEN
Ling WANG
Ling CHEN
Ling1 WANG
Liqing ZHANG
Liquan SHEN
Lixin ZHENG
Luo ZHONG
Lusheng ZHONG
Luyang GUAN
Manjaiah D H
Maoguo GONG
Maoguo GONG
Maoyuan ZHANG
Masahiko TOYONAGA
Masakazu MATSUGU
Masumi ISHIKAWA
Mehmet Zeki BILGIN
Mei TIAN
Meihong SHI
Meiyi LI
Mengxin LI
Michael MARGALIOT
Min LIU
Min FANG
Ming BAO
Ming LI
Ming LI
Ming CHEN
Mingbao LI
Mingguang WU
Minghui LI
Mingquan ZHOU
Moh Lim SIM
Mudar SAREM
Nagabhushan P.
Naigang CUI
Nak Yong KO
Naoko TAKAYAMA
Naoyuki KUBOTA
Ning CHEN
Otvio Noura TEIXEIRA
Pei-Chann CHANG
Peide LIU
Peixin YE
Peizhi WEN

Peng TIAN
Peter TINO
Phill Kyu RHEE
Ping JI
Pu WANG
Qi WANG
Qi LUO
Qiang LV
Qiang SUN
Qijuan CHEN
Qing GUO
Qing LI
Qinghe MING
Qingming YI
Qingqi PEI
Qiongshui WU
Qiyong GONG
Quan ZHANG
Renbiao WU
Renpu LI
Renren LIU
Richard EPSTEIN
Richard NEVILLE
Robo ZHANG
Roman NERUDA
Rong LUO
Rongfang BIE
Ronghua SHANG
Ronghua SHANG
Rubin WANG
Rui XU
Ruijun ZHU
Ruiming FANG
Ruixuan LI
Ruochen LIU
S.G. LEE
Sanyang LIU
Satoshi MATSUDA
Seok-Lyong LEE
Seong Whan KIM
Serdar KUCUK
Seunggwan LEE
Sezai TOKAT
Shan TAN
Shangmin LUAN
Shao-Ming FEI
Shao-Xiong WU
Shigeo ABE
Shiqiang ZHENG
Shuguang ZHAO
Shuiping GOU
Shui-Sen CHEN
Shui-Sheng ZHOU
Shunman WANG
Shunsheng GUO
Shutao LI
Shuyuan YANG
Soo-Hong PARK
Soon Cheol PARK
Sung-Bae CHO
Sungshin KIM
Sunjun LIU
Sunkook YOO
Tae Ho CHO
Tae-Chon AHN
Tai Hoon CHO
Takao TERANO
Takeshi FURUHASHI
Tan LIU
Tao SHEN
Tao WANG
Taoshen LI
Thi Ngoc Yen PHAM
Tianding CHEN
Tiantian SU
Tianyun CHEN
Tie-Jun ZHOU
Ting WU
Tong-Zhu FANG
Vianey Guadalupe CRUZ SANCHEZ
Vic CIESIELSKI
Wang LEI
Wanli MA
Wei ZOU
Wei WU
Wei LI
Wei FANG
Weida ZHOU
Wei-Hua LI
Weiqin YIN

Weiyou CAI
Wei-Yu YU
Wen ZHU
Wenbing XIAO
Wenbo XU
Wenchuan YANG
Wenhui LI
Wenping MA
Wenping MA
Wenqing ZHAO
Wen-Shyong TZOU
Wentao HUANG
Wentao HUANG
Wenxing ZHU
Wenxue HONG
Wenyu LIU
X.B. CAO
Xian-Chuan YU
Xianghui LIU
Xiangrong ZHANG
Xiangwei LAI
Xiaobing LIU
Xiaodong KONG
Xiaofeng SONG
Xiaoguang ZHANG
Xiaoguang LIU
Xiaohe LI
Xiaohua YANG
Xiaohua WANG
Xiaohua ZHANG
Xiaohui YUAN
Xiaohui YANG
Xiaojian SHAO
Xiao-Jie ZHAO
Xiaojun WU
Xiaoli LI
Xiaosi ZHAN
Xiaosuo LU
Xiaoyi FENG
Xiaoying PAN
Xiaoyuan WANG
Xin XU
Xin YUAN
Xinbo GAO
Xinchao ZHAO
Xingming SUN
Xinsheng YAO
Xinyu WANG
Xiu JIN
Xiu-Fen YU
Xiufeng WANG
Xiuhua GAO
Xiuli MA
Xiyang LIU
Xiyue HUANG
Xu YANG
Xu CHEN
Xuejun XU
Xueliang BI
Xuerong CHEN
Xuezhou XU
Xun WANG
Xuyan TU
Yan ZHANG
Yan LIANG
Yan ZHANG
Yang YAN
Yangmi LIM
Yangmin LI
Yangyang LI
Yangyang WU
Yanling WU
Yanning ZHANG
Yanning ZHANG
Yanpeng LIU
Yanping LV
Yanxia ZHANG
Yanxin ZHANG
Yan-Xin ZHANG
Yaoguo DANG
Yaping DAI
Yaw-Jen CHANG
Yeon-Pun CHANG
Yezheng LIU
Yidan SU
Yifeng NIU
Yimin YU
Ying GUO
Ying GAO
Ying TIAN

Yingfang FAN
Yingfeng QIU
Yinghong PENG
Yingying LIU
Yong ZHAO
Yong YANG
Yong FAN
Yong-Chang JIAO
Yonggui KAO
Yonghui JIANG
Yong-Kab KIM
Yongqiang ZHANG
Yongsheng DING
Yongsheng ZHAO
Yongzhong ZHAO
Yoshikii KASHIMORI
You-Feng LI
Youguo PI
You-Ren WANG
Yu GUO
Yu GAO
Yuan KANG
Yuehui CHEN
Yuehui CHEN
Yufeng LIAO
Yuheng SHA
Yukun BAO
Yulong LEI
Yumin LIU
Yumin TIAN
Yun-Chia LIANG
Yunjie ZHANG
Yuping WANG
Yurong ZENG
Yusuke NOJIMA
Yutao QI
Yutian LIU
Yuyao HE
Yu-Yen OU
Yuzhong CHEN
Zafer BINGUL
Zeng-Guang HOU
Zhang YANG
Zhanli LI
Zhao ZHAO
Zhaoyang ZHANG
Zhe-Ming LU
Zhen YANG
Zhenbing ZENG
Zhengxing CHENG
Zhengyou XIA
Zhi LIU
Zhidong ZHAO
Zhifeng HAO
Zhigang XU
Zhigeng FANG
Zhihui LI
Zhiqing MENG
Zhixiong LIU
Zhiyong ZHANG
Zhiyu ZHANG
Zhonghua LI
Zhurong WANG
Zi-Ang LV
Zixing CAI
Zong Woo GEEM
Zongmin LI
Zongying OU
Zoran BOJKOVIC

Table of Contents – Part II

Other Topics in Natural Computation

Natural Computation Techniques Applications

Hardware

Cross-Disciplinary Topics

Table of Contents – Part I

Artificial Neural Networks

Natural Neural Systems and Cognitive Science

Neural Network Applications

Evolutionary Computation: Theory and Algorithms

Simulation and Investigation of Quantum Search Algorithm System

Sun Li[1] and Xu Wen-Bo[2]

[1] School of E-learning, Southern Yangtze University, Wuxi, P.R. China 214036
lisun@sytu.edu.cn
[2] School of Information Engineer, Southern Yangtze University, Wuxi, P.R. China 214122
xwb@sytu.edu.cn

Abstract. NMR has been considered as one of the most effective physical system to realize quantum computation. As multiple-quantum operator algebra theory mentioned, any unitary transformation can be decomposed into a sequence of a limited number of one-qubit quantum gates and two-qubit diagonal gates. Based on these backgrounds, we proposed the method to design NMR pulse sequences to implement Generalized Quantum Search Algorithm with arbitrary phase rotation, and, experimental finished the algorithm with different phase rotations respectively in a two-qubit system, on a Quantum Computer Emulator.

Keywords: NMR, Generalized Quantum Search Algorithm, phase rotations, NMR pulse sequences.

1 Introduction

Grover's Quantum Search Algorithm (Grover, 1996) is optimal. Depending on the superiority of quantum parallel computation, the algorithm has decreased query number to search an item in an unsorted database squarely, compared to classical search algorithms. As a next wave of evolution, building blocks of the algorithm are aimed to be defined as general as possible. The original Grover algorithm has been enhanced to Generalized Quantum Search Algorithm (Grover, 1998).

Quantum computation follows the laws of quantum mechanics. The two-state quantum system which represents qubit should be interacted. The interaction can be used in computation. So the computation would be underway though a special extra action, which operates and controls the change of the state. Besides this, the system should be independence. This ensures the coherence of quantum states not to be disrupted. Because of the weak interaction with outsides, spin-1/2 for every nuclear (up and down), and the spin coupling between two spins, together with the extra radio frequency (RF) pulses to control and survey the spin, Nuclear Magnetic Resonance (NMR) system is coming to be one of the mostly used physical systems to implement quantum computation.

To implement quantum computation through NMR, quantum gates should be turned into NMR pulse sequences. As multiple-quantum operator algebra theory

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 1 – 9, 2006.

mentioned (Miao, 2000), any unitary transformation can be decomposed into a sequence of a limited number of one-qubit quantum gates and two-qubit diagonal gates. In this paper, we decomposed quantum gates in Generalized Quantum Search Algorithm in a 2-qubit system with different phase rotating angles, designed NMR pulse sequences to realize those gates. Finally, we verified the design with a Quantum computation Emulator (Michielsen et al., 2003).

2 Generalized Quantum Search Algorithm

Beginning with an average probability amplitude superposition state to be the initial state, Grover algorithm consists of four steps in a iteration: (1) a phase inversion of the target state $|s\rangle$, with the operator $I_s=I-2|s\rangle\langle s|$; (2) the Walsh–Hadamard transformation W; (3) a phase inversion of the prepared state $|00\ldots0\rangle$, with the operator $I_0=I-2|0\rangle\langle 0|$ and (4) the Walsh–Hadamard transformation W. Step 2–4 are combined to give the inversion about average $D=WI_0W$. The operator for one Grover iteration is $Q=WI_0WI_s$. After $T=(\pi/4)\sqrt{N}$ iterations, the probability amplitude of the marked state $|s\rangle$ is nearly 1.

In Generalized Quantum Search Algorithm, phase inversions are replaced by arbitrary phase rotations. An iteration is composed of four steps as follows: (1) a phase rotation of the target state $|s\rangle$ with the angle θ, the operator is $I^{g}_{s}=I-(1-e^{i\theta})|s\rangle\langle s|$; (2) a arbitrary unitary transformation U; (3) a phase rotation of the prepared state $|00\ldots0\rangle$, with the angle φ, the operator is $I^{g}_{0}=I-(1-e^{i\varphi})|0\rangle\langle 0|$ and (4) an arbitrary unitary transformation U.

Obviously, when U turns into W and $\theta=\varphi=\pi$, Grover algorithm is recovered. Further research indicated that only when phase matching was satisfied, which is $\theta=\varphi$, that Generalized Quantum Search Algorithm would work (Long et al., 1999). With different phase rotating angle θ (φ), the iteration number is not the same.

3 Design of NMR Pulse Sequences for Generalized Quantum Search Algorithm

3.1 Multiple-Quantum Operator Algebra Theory

Quantum computation is a reversible process. It can be thought of as a series of unitary transformations acted on the input state, that is: $|\varphi(t)=\prod_i U_i(t)|\varphi_0\rangle$.

The time-evolutional operator $U(t)$ obeys the basic law of quantum mechanics-Schrödinger Equation:

$$\frac{dU(t)}{dt}=-iH(t)U(t) \qquad (\hbar=1) \tag{1}$$

where $H(t)$ is the effective Hamiltonian for the unitary operator $U(t)$. Then it follows from (1) that:

$$U(t) = \exp[-iH(t)t] \tag{2}$$

The research of NMR (Chuang et al., 1998) mentioned that in a strong static multi-dimensional magnetic field, the spin Hamiltonian could be expressed as:

$$H_0 = h_0 + \sum_{k=1}^{N} h_k' I_{kz} + \sum_{l>k=1}^{N} 2J_{kl} I_{kz} I_{lz} + \sum_{m>l>k=1}^{N} 4J_{klm} I_{kz} I_{lz} I_{mz} + ... \tag{3}$$

where h_k' and I_{kz} are planck parameter and spin angular momentum operator in Z direction for spin k respectively. J is the coefficient of spin-coupling.

Following from (2) and (3), we gain the operator U_0 in a multi-qubit quantum computation system that:

$$U_0 = \exp(-ih_0)\prod_{k=1} \exp(-ih_k' I_{kz}) \prod_{l>k=1} \exp(-i2J_{kl} I_{kz} I_{lz}) \prod_{m>l>k=1} \exp(-i4J_{mlk} I_{kz} I_{lz} I_{mz})... \tag{4}$$

where *exp(-ih$_0$)* and other constant items can be ignored. (4) can be further decomposed into production of a series of single-qubit gates and two-qubit CNOT gates. In a two-spin NMR system, as a two-qubit system, unitary operator U_0 can be expressed as that:

$$U_0 = \exp(-ih_1' I_{1z}) \exp(-ih_2' I_{2z}) \exp(-iJ_{12} I_{1z} I_{2z}) \tag{5}$$

3.2 Elementary Quantum Gate, Walsh-Hadamard Gate and Spin-Coupling Item

Those spin angular momentum operators in a spin-1/2 system can be denoted in the form of the famous Pauli Matrix in Atomic Physics, which is:

$$I_x = \frac{1}{2}\sigma_x = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ I_y = \frac{1}{2}\sigma_y = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},\ I_x = \frac{1}{2}\sigma_x = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

The operator making the spin to rotate in an angle θ around the x-axis, which requires a pulse along the *y*-axis, can be defined as $X^{\theta} = \exp[i\theta I_x]$. $\overline{X}^{\theta}$ means an inverse rotation to X^{θ}, $\overline{X}^{\theta} = \exp[-i\theta I_x]$.

Of particular interest, when $\theta=\pi/2$,

$$X^{\pi/2} = \exp[i\pi I_x / 2] = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$$

$$\overline{X}^{\pi/2} = \exp[-i\pi I_x / 2] = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix}$$

It is easily to give that: $X^{\pi/2} |0> = 1/\sqrt{2}(|0> + i|1>)$,

$$\overline{X}^{\pi/2} |0> = 1/\sqrt{2}(i|0> + |1>).$$

So $X^{\pi/2}$ and $\overline{X}^{\pi/2}$ can be taken as elementary quantum gates.

Same to the rotation around y-axis, which needs a pulse along x-axis.

$$Y^{\theta} = \exp[i\theta I_y], \ \overline{Y}^{\theta} = \exp[-i\theta I_y].$$

$$Y^{\pi/2} = \exp[i\pi I_y / 2] = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix},$$

$$\overline{Y}^{\pi/2} = \exp[-i\pi I_y / 2] = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

$$Y^{\pi/2} \mid 0>= 1/\sqrt{2}(\mid 0> - \mid 1>), \ \overline{Y}^{\pi/2} \mid 0>= 1/\sqrt{2}(\mid 0> + \mid 1>).$$

In a NMR system, the RF pulse can only be lead into along x or y-axis, so the spin operator making a rotation around the z-axis can be denoted as a combination of X^{θ} and Y^{θ}, which is $Z^{\theta} = Y^{\frac{\pi}{2}} X^{\theta} \overline{Y}^{\frac{\pi}{2}}$, $\overline{Z}^{\theta} = Y^{\frac{\pi}{2}} \overline{X}^{\theta} \overline{Y}^{\frac{\pi}{2}}$.

$$Z^{\pi/2} \mid 0>= \exp[i\pi/4] \mid 0>, \ \overline{Z}^{\pi/2} \mid 0>= \exp[-i\pi/4] \mid 0>.$$

$$Z^{\pi/2} \mid 1>= \exp[-i\pi/4] \mid 1>, \ \overline{Z}^{\pi/2} \mid 1>= \exp[i\pi/4] \mid 1>.$$

Another basic operation is the single-qubit Walsh-Hadamard transform W, which, in terms of elementary rotations, reads

$$\mathrm{W} = \frac{i}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = (X^{\frac{\pi}{2}})^2 \overline{Y}^{\frac{\pi}{2}} = Y^{\frac{\pi}{2}} (X^{\frac{\pi}{2}})^2 \tag{6}$$

Multi-qubit W-H transform can be expressed as an orderly conduction by a series of single-qubit W-H transforms.

$$W^n = W_1 \otimes W_2 \otimes ... \otimes W_n \tag{7}$$

The spin-coupling item in a two-qubit system is defined as:

$$G = \exp[-2\pi i J_{12} I_{1z} I_{2z} t], \text{ in which: } t = (2\pi - \theta)/2J_{12}\pi.$$

It gives: $$G = \exp[-i(2\pi - \theta) I_{1z} I_{2z}] \tag{8}$$

3.3 NMR Pulse Sequences for Generalized Quantum Search Algorithm

On the basis of (5)-(8), we give NMR pulse sequences for all operators in a two-qubit Generalized Quantum Search Algorithm.

W-H Transform: $W^2 = (X_1^{\frac{\pi}{2}})^2 \overline{Y}_1^{\frac{\pi}{2}} (X_2^{\frac{\pi}{2}})^2 \overline{Y}_2^{\frac{\pi}{2}}$ (9)

Phase inversion operator for |*00*>: $I_0^{\,g} = Y_1^{\frac{\pi}{2}} \overline{X}_1^{\frac{\theta}{2}} \overline{Y}_1^{\frac{\pi}{2}} Y_2^{\frac{\pi}{2}} \overline{X}_2^{\frac{\theta}{2}} \overline{Y}_2^{\frac{\pi}{2}} G$ (10)

Phase inversion operator for |*10*>: $I_1^{\,g} = Y_1^{\frac{\pi}{2}} X_1^{\frac{\theta}{2}} \overline{Y}_1^{\frac{\pi}{2}} Y_2^{\frac{\pi}{2}} \overline{X}_2^{\frac{\theta}{2}} \overline{Y}_2^{\frac{\pi}{2}} G$ (11)

Phase inversion operator for |*01*>: $I_2^{\,g} = Y_1^{\frac{\pi}{2}} \overline{X}_{1\frac{\theta}{2}} \overline{Y}_{1\frac{\pi}{2}} Y_2^{\frac{\pi}{2}} X_2^{\frac{\theta}{2}} \overline{Y}_{2\frac{\pi}{2}} G$ (12)

Phase inversion operator for |*11*>: $I_3^{\,g} = Y_1^{\frac{\pi}{2}} X_1^{\frac{\theta}{2}} \overline{Y}_{1\frac{\pi}{2}} Y_2^{\frac{\pi}{2}} X_2^{\frac{\theta}{2}} \overline{Y}_{2\frac{\pi}{2}} G$ (13)

The sequence to finish the whole search is,

$$Q_j^{\,g} \mid \phi_0 >= (W^2 I_0^{\,g} W^2 I_j^{\,g})^T W^2 \mid \phi_0 > \tag{14}$$

Where j is the target for search, $|\Phi_0>$ is the prepared state |00> and T is the iteration number.

4 Simulating Implementation and Results

Quantum Computer Emulator (QCE) is a classic computer program with graphical user interface. It emulates various hardware designs of Quantum Computer, simulates the physical processes that govern the operation of the hardware quantum processor, strictly according to the laws of quantum mechanics. The QCE also provides an ideal NMR environment to debug and execute quantum algorithms under realistic experimental conditions. In QCE, all operations to a spin, which require extra RF pulse, can be achieved through Micro-Instruction (MI). A series of MIs compose quantum program to constitute quantum gates and the complete quantum algorithm. The key to design a MI is the definition of some significant parameters, such like operating duration $\tau/2\pi$, J and h.

We designed QCE Micro-Instructions required to implement generalized search algorithm in a 2-qubit NMR system under conditions described below.

In an ideal 2-spin NMR system, any operation to a spin by a RF pulse can be denoted as, $U(t) = \exp(-iH\tau)$ (15)

$$H = -JI_{1z}I_{2z} - h_1^{\,x} I_{1x} - h_2^{\,x} I_{2x} - h_1^{\,y} I_{1y} - h_2^{\,y} I_{2y} - h_1^{\,z} I_{1z} - h_2^{\,z} I_{2z} \tag{16}$$

When we considerate the operator $X_j^{\,\theta} = \exp[i\theta I_{j_x}]$, $H = -h_j^{\,x} I_j^{\,x}$, it gives $ih_j^{\,x} I_{j_x}\tau = i\theta I_{j_x}$ (j=1,2). So the MI to achieve the operator $X_j^{\,\theta}$ meets:

$$\tau / 2\pi = \theta / 2\pi h_j^{\,x} \tag{17}$$

Same to the MI for operator $Y_j^{\,\theta}$, $\tau / 2\pi = \theta / 2\pi h_j^{\,y}$ (18)

To spin-coupling item: $G = \exp[-i(2\pi - \theta)I_{1z}I_{2z}]$, $H = -JI_{1z}I_{2z}$, it gives

$$\tau / 2\pi = (2\pi - \theta) / 2\pi J \tag{19}$$

In the simulation, we implemented Generalized Quantum Search Algorithm with three different phase rotation, π, 3π/4 and π/2 respectively. All significant parameters for MI design required in the experiment, according to (17)-(19), are shown in table 1.

Table 1. QCE MI parameters in the simulation

MI	τ/2π	H_1^x	H_2^x	H_1^y	H_2^y	H_1^z	H_2^z	J_{12}^z
$X_1(\pi/2)$	0.25	1	0	0	0	0	0	0
$X_2(\pi/2)$	0.25	0	1	0	0	0	0	0
$X_1(-\pi/2)$	0.25	-1	0	0	0	0	0	0
$X_2(-\pi/2)$	0.25	0	-1	0	0	0	0	0
$Y_1(\pi/2)$	0.25	0	0	1	0	0	0	0
$Y_2(\pi/2)$	0.25	0	0	0	1	0	0	0
$Y_1(-\pi/2)$	0.25	0	0	-1	0	0	0	0
$Y_2(-\pi/2)$	0.25	0	0	0	-1	0	0	0
$X_1(3\pi/8)$	0.1875	1	0	0	0	0	0	0
$X_2(3\pi/8)$	0.1875	0	1	0	0	0	0	0
$X_1(-3\pi/8)$	0.1875	-1	0	0	0	0	0	0
$X_2(-3\pi/8)$	0.1875	0	-1	0	0	0	0	0
$X_1(\pi/4)$	0.125	1	0	0	0	0	0	0
$X_2(\pi/4)$	0.125	0	1	0	0	0	0	0
$X_1(-\pi/4)$	0.125	-1	0	0	0	0	0	0
$X_2(-\pi/4)$	0.125	0	-1	0	0	0	0	0
$G(\pi)$	$5*10^5$	0	0	0	0	0	0	$-1*10^{-6}$
$G(3\pi/4)$	$6.25*10^5$	0	0	0	0	0	0	$-1*10^{-6}$
$G(\pi/2)$	$7.5*10^5$	0	0	0	0	0	0	$-1*10^{-6}$

τ/2π corresponds to the QCE MI parameter 'Time step'. We set 'Main steps' and 'Intermediate step' to 1. All other parameters not in table 1 are set to 0.

After finishing Micro-Instructions, we composed all quantum programs with three phase rotations based on (9) to (14), and formed iterations for quantum search algorithm. Fig. 1 to 3 gave the QCE results to implement generalized quantum search to reach single target state $|00\rangle$, $|10\rangle$, $|01\rangle$ and $|11\rangle$, with different phase rotation, π, 3π/4 and π/2 respectively. The QCE results are, when phase rotation is π, which means classic Grover search algorithm, only one iteration would lead to the target state; 3 iterations are needed when phase rotating 3π/4; iteration number is 6 with π/2 phase rotation. The result in π/2 is in accordance with the conclusion to search for $|11\rangle$ in a concrete NMR system reported by Long GL (Long et al., 2001).

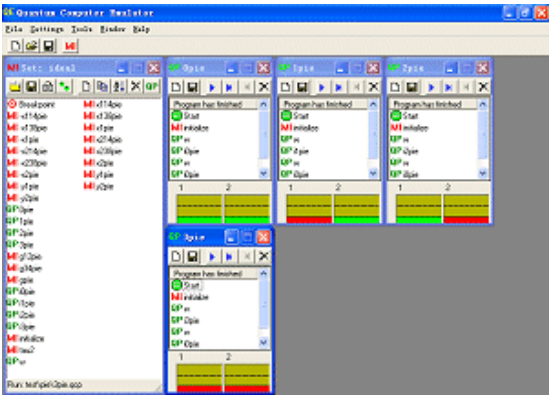


Fig. 1. QCE result (phase rotation =π)

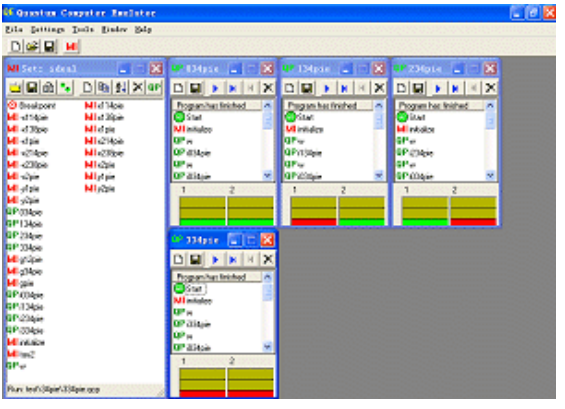


Fig. 2. QCE result (phase rotation =3π/4)

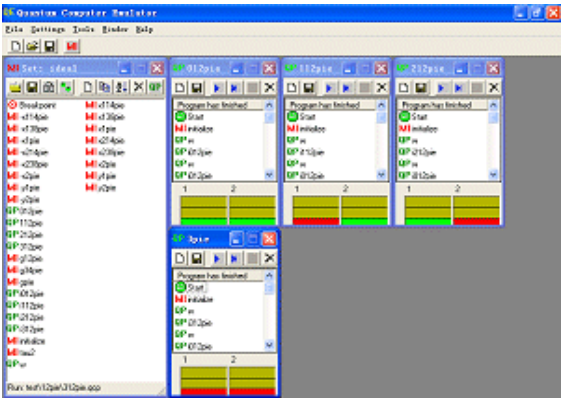


Fig. 3. QCE result (phase rotation =π/2)

5 Conclusion

The simulating results stated that those NMR pulse sequences we designed to implement Generalized Quantum Search Algorithm were on the whole reasonable. Although Generalized Quantum Search Algorithm has extended Grover quantum search algorithm effectively, those integrated expenses to practically implement the search is not optimal. That means, phase inversion is the most economical method to finish the iteration in the search step, even though it is not easy to control the spin rotation in real NMR system.

An important topic that we did not touch upon at all is the effect of the interaction of the NHR with its environment (dissipation, decoherence). Dissipation can not be treated within the context of the models that the QCE can solve. We need to solve the equations of motion of the full density matrix, instead of solving the time-dependent Schrödinger equation, if we take into account of the dissipation. On the other hand, decoherence can be studied within QCE, though we did not include in the paper. Not considering the interaction may bring about some affections to the identification of simulating effects, but we think they would not be strong enough to change the iteration number of the quantum search.

Due to the easy use of QCE on a personal computer, its exact simulation to the real NHR environment, and NHR being regarded as one of the mostly used physical systems to implement quantum computation, we think it show advantages compared to other simulations of quantum search algorithm, which not being realized on a PC or in a NHR environment (Jumpei Niwa, Keiji Matsumoto, Hiroshi Imai, 2002 and Shingo Fujiwara, Shuichi Hasegawa, 2005). A computer simulation of NMR implementing quantum algorithm provides certain reference senses to theoretical research of quantum computation and algorithm, verification for feasibility of quantum algorithm, as well as physical achievement of quantum computer.

References

1. Grover LK. A Fast Quantum Mechanical Algorithm for Database Search[C]. In: Proc of the 28th Annual ACM Symposium on Theory of Computing, 1996: 212-219.
2. Grover LK. Quantum computers can search rapidly by using almost any transformation[J]. Phys. Rev. Lett. A 80, 1998: 4329-4332.
3. Miao X. Universal Construction of Unitary Transformation of Quantum Computation with One- and Two-body Interactions[J]. http://xxx.lanl.gov/abs/quant-ph/0003068.
4. Michielsen K, Raedt HD. QCE: A Simulator for Quantum Computer Hardware[J]. Turk J. Phys., 2003, 27: 129.
5. http://rugth30.phys.rug.nl/compphys.
6. Long GL, Li YS, Zhang WL, et al. Phase matching in quantum searching [J]. Phys Lett A, 1999, A262: 27-34.
7. Chuang IL, Gershenfeld NA, et al. Bulk quantum computation with nuclear magnetic resonance: theory and experiment [J]. Proc R Soc Load A, 1998, 454: 447-467.

8. Long GL, Yan HY, Li YS, et al. Experimental NMR realization of a Generalized Quantum Search Algorithm [J]. Phys Lett A, 2001, A286: 121-126.
9. Jumpei Niwa, Keiji Matsumoto, Hiroshi Imai. General-purpose parallel simulator for quantum computing[J]. Phys. Rev. A, 66, 2002, 062317.
10. Shingo Fujiwara, Shuichi Hasegawa. General method for realizing the conditional phase-shift gate and a simulation of Grover's algorithm in an ion-trap system[J]. Phys. Rev. A, 71, 2005, 012337.

Quantum Integration Error for Some Sobolev Classes

Ye Peixin[1,⋆] and Hu Xiaofei[2]

[1] School of Mathematical Sciences and LPMC, Nankai University, Tianjin 300071, China
yepx@nankai.edu.cn
[2] College of Mathematical, Syracuse University, NY 13210, USA

Abstract. Based on the quantum algorithms for approximating the mean of p-summable sequences, we develop quantum algorithms for approximating the integration from anisotropic and generalized Sobolev classes. Then we analyze query error of these algorithms and prove their optimality. It is proved that for these classes of functions the optimal convergence rates of quantum algorithms are essential smaller than those of classical deterministic and randomized algorithms. Our results extend recent works of Heinrich on classical Sobolev classes.

1 Introduction and Results

An important question in the overlap of computer science, mathematics, and physics, is the exploration of potential capabilities of quantum computers. Milestones work is Shor's discovery of polynomial factoring algorithm on a quantum computer and the quantum search algorithm of Grover, cf. [7, 10]. So far, major research was concentrated on discrete problems that one encounters in computer science. While for the quantum computation of numerical problem, the pioneering work is the establishment of the matching upper and lower bound for the numerical integration from classical Hölder classes which was done by Novak, cf. [9]. After that, a series of papers about summation of p-summable sequences and multivariate integration of various function classes were published, cf. [3–6] and the references therein. In this paper, we extend the work of Heinrichs' about classical Sobolev classes by studying the optimal order of quantum integration error for anisotropic Sobolev classes and generalized Sobolev classes. Comparing our results with the exact order of these classes in the classical deterministic and randomized settings cf. [2, 11], we see that quantum algorithm bring a essential speedups over classical algorithms.

Let's recall the quantum computation model for numerical problems. For nonempty set Ω and K we denote the set of all function from Ω to K by $\mathcal{F}(\Omega, K)$. Let G be a normed space with scalar field $\mathbb{K}$, which is either $\mathbb{R}$ or $\mathbb{C}$, and let S be any mapping from F to G, where $F \subset \mathcal{F}(\Omega, \mathbb{R})$. We want to approximate $S(f)$ for $f \in F$ by quantum computations. Denote $\mathbb{Z}[0, N) := \{0, \ldots, N-1\}$ for

⋆ Corresponding author, supported by the Natural Science Foundation of China (Grant No. 10501026).

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 10–19, 2006.

$N \in \mathbb{N}$. Let H_m be m-fold tensor product of H_1, two-dimensional Hilbert space over $\mathbb{C}$, and let $\{e_0, e_1\}$ be two orthonormal basis of H_1. An orthonormal basis of H_m, denoted by $\mathcal{C}_m$, consist of the vectors $|l>:= e_{i_0} \otimes \ldots \otimes e_{i_m}$ $(l \in \mathbb{Z}[0, 2^{m-1}))$, where $\otimes$ is the tensor product, $i_j \in \{0, 1\}$ and $l = \sum_{j=0}^{2^m-1} i_j 2^{m-1-j}$. Let $\mathcal{U}(H_m)$ stand for the set of unitary operator on H_m. Two mappings is defined respectively by $\tau : Z \to \Omega$ and $\beta : K \to \mathbb{Z}[0, 2^{m''})$. where for m, m', m'' $\in \mathbb{N}$, $m' + m'' \leq m$ and Z is the nonempty subset of $\mathbb{Z}[0, 2^{m'})$. A quantum query on F is given by a tuple $Q = (m, m', m'', Z, \tau, \beta)$, and the number of quits $m(Q) := m$. We define the unitary operator Q_f for a given query Q by setting for each $f \in F$

$$Q_f|i> |x> |y> := \begin{cases} |i> |x \oplus \beta(f(\tau(i)))> |y> & \text{if } i \in Z, \\ |i> |x> |y> & \text{otherwise,} \end{cases} \tag{1}$$

where set $|i> |x> |y> \in \mathcal{C}_m := \mathcal{C}_{m'} \otimes \mathcal{C}_{m''} \otimes \mathcal{C}_{m-m'-m''}$ and denote addition modulo $2^{m''}$ by $\oplus$. Let tuple $A = (Q, (U_j)_{j=0}^n)$ denote a quantum algorithm on F with no measurement, where Q is a quantum query on F, $n \in \mathbb{N}_0$ ($\mathbb{N}_0 = \mathbb{N} \bigcup \{0\}$) and $U_j \in \mathcal{U}(H_m)$, with $m = m(Q)$. For each $f \in F$, we have $A_f \in \mathcal{U}(H_m)$ with the following form

$$A_f = U_n Q_f U_{n-1} \ldots U_1 Q_f U_0. \tag{2}$$

The number of queries are denoted by $n_q(A) := n$. A quantum algorithm $A : F \to G$ with k measurements is defined by for $k \in \mathbb{N}$

$$A = ((A_l)_{l=0}^{k-1}, (b_l)_{l=0}^{k-1}, \phi), \tag{3}$$

where $A_l (l \in \mathbb{Z}[0, k))$ are quantum algorithms on F with no measurements, $b_0 \in \mathbb{Z}[0, 2^{m_0})$ is a fixed basis state with which A starts. For $1 \leq l \leq k-1$, apply the quantum operations to b_{l-1} and get a random state ζ_{l-1}. The resulting state ζ_{l-1} is memorized and transformed into a new basis state b_l,

$$b_l : \prod_{i=0}^{l-1} \mathbb{Z}[0, 2^{m_i}) \to \mathbb{Z}[0, 2^{m_l}), \tag{4}$$

where we denote $m_l := m(A_l)$ and ϕ is a function

$$\phi : \prod_{i=0}^{k-1} \mathbb{Z}[0, 2^{m_l}) \to G. \tag{5}$$

Let $n_q := \sum_{l=0}^{k-1} n_q(A_l)$ denote the number of queries used by A and $(A_f(x, y))_{x,y \in \mathcal{C}_m}$ the matrix of the transformation A_f in the canonical basis $\mathcal{C}_m$, $A_f(x, y) =< x|A_f|y >$. The output of A at input $f \in F$ will be a probability measure $A(f)$ on G, defined as follows:

$$\begin{aligned} p_{A,f}(x_0, \ldots, x_{k-1}) = {} & |A_{0,f}(x_0, b_0)|^2 |A_{1,f}(x_1, b_1(x_0))|^2 \\ & \ldots |A_{k-1,f}(x_{k-1}, b_{k-1}(x_0, \ldots, x_{k-2}))|^2. \end{aligned} \tag{6}$$

Define $A(f)$ by

$$A(f)(C) = \sum_{\phi(x_0,\ldots,x_{k-1})\in C} p_{A,f}(x_0,\ldots,x_{k-1}), \qquad \forall C \subset G. \tag{7}$$

The error of A for S on input f is defined as follows: Let $0 \le \theta < 1$, $f \in F$, let ξ be any random variable with distribution $A(f)$, and let $e(S, A, f, \theta) = \inf\{\epsilon \ge 0 : P\{\|s(f) - \xi\| > \epsilon\} \le \theta\}$. Associated with this we introduce

$$e(S, A, F, \theta) = \sup_{f\in F} e(S, A, f, \theta), \tag{8}$$

$$e(S, A, f) = e(S, A, f, \frac{1}{4}), \tag{9}$$

$$e(S, A, F) = \sup_{f\in F} e(S, A, f). \tag{10}$$

The n-th minimal query error is defined for $n \in \mathbb{N}_0$ as

$$e_n^q(S, F) := \inf\{e(S, A, F) : \quad n_q(A) \le n\}. \tag{11}$$

Let $D = [0, 1]^d$, $C(D)$ be the space of continuous functions on D with the supremum norm. For $1 \le p < \infty$, let $L_p(D)$ be the space of p-th Lesbegue-integrable functions with the usual norm. Below for any Banach space X the unit ball centered at the origin is denoted by $\mathcal{B}(X)$, which is defined as $\{f \in X : \|f\|_X \le 1\}$.

Let $\mathbf{e}_j = (\delta_{i,j})$, the Kronecker notation. For $\mathbf{r} \in \mathbb{N}^d$ and $1 \le p < \infty$, the anisotropic Sobolev space $W_p^{\mathbf{r}}(D)$ consists of functions $f \in L_p(D)$ such that for $j \in \{1, ..., d\}$, $\partial^{r_j \mathbf{e}_j} f \in L_p(D)$. It is known that the space $W_p^{\mathbf{r}}(D)$ is a Banach space with the norm

$$\|f\|_{W_p^{\mathbf{r}}(D)} := \|f\|_{L_p(D)} + \sum_{j=1}^{d} \|\partial^{r_j \mathbf{e}_j} f\|_{L_p(D)}. \tag{12}$$

We use the notation $g(\mathbf{r}) = (\sum_{j=1}^{d} \frac{1}{r_j})^{-1}$ and assume that $g(\mathbf{r}) > 1/p$, so that the space $W_p^{\mathbf{r}}(D)$ can be imbedded into $C(D)$. Define the integration operator $I_d : W_p^{\mathbf{r}}(D) \to \mathbb{R}$ by

$$I_d(f) = \int_D f(\mathbf{t}) d\mathbf{t}. \tag{13}$$

Theorem 1. *Let $\mathbf{r} \in \mathbb{N}^d$, $1 \le p < \infty$ and assume $g(\mathbf{r}) > 1/p$, then*

$$n^{-g(\mathbf{r})-1} \le e_n^q(I_d, \mathcal{B}(W_p^{\mathbf{r}}(D))) \le c \cdot n^{-g(\mathbf{r})-1} \cdot \alpha_p, \tag{14}$$

where

$$\alpha_p := \begin{cases} 1 & \text{if } 2 < p \le \infty, \\ (\log\log n)^{3/2} \log\log\log n & \text{if } p = 2, \\ (\log n)^{2/p-1} & \text{if } 1 \le p < 2, \end{cases} \tag{15}$$

Note that the results of the case $r_1 = r_2 = \cdots = r_d$ were obtained by Heinrich. Then we consider more general Sobolev class. A set $\Lambda \subset \mathbb{N}_0^d$ is said to be complete, if it is bounded and satisfies: $\alpha \in \Lambda$ and $\alpha' \le \alpha$, implies that $\alpha' \in \Lambda$, and the boundary of Λ is defined by

$$\partial\Lambda := \{\, \alpha : \ \alpha \notin \Lambda, \text{ and if } \alpha' < \alpha, \text{ then } \alpha' \in \Lambda \,\}, \tag{16}$$

The generalized Sobolev class is defined as

$$\mathcal{B}(W_p^{\Lambda}(D)) := \big\{ f \in C(D) : \sum_{\alpha \in \partial\Lambda} \|D^{\alpha} f\|_{L_p(D)} + \|f\|_{L_p(D)} \le 1 \big\}. \tag{17}$$

Theorem 2. *Let Λ be a complete set, $1 \le p < \infty$, then*

$$e_n^q(I_d, \mathcal{B}(W_p^{\Lambda}(D))) \le c \cdot n^{-g(\mathbf{r}_{\Lambda})-1} \cdot \alpha_p \,, \tag{18}$$

$\mathbf{r}_{\Lambda} = (r_1, \ldots, r_d)$ satisfying $r_j \mathbf{e}_j \in \partial\Lambda$ for $j = 1, \ldots, d$.

It is known from [2] and [11] that the optimal bound of integration error for the classes $\mathcal{B}(W_p^{\mathbf{r}}(D))$ is $n^{-g(\mathbf{r})}$ while that of randomized algorithms is $n^{-g(\mathbf{r})-\min\{1-1/p,\,1/2\}}$. As the classical Sobolev classes we find that for there is a e essential speedup of quantum algorithm over classical deterministic and quadratic speedup over randomized algorithm. Similar comments can be made for the class $\mathcal{B}(W_p^{\Lambda}(D))$, cf. [2].

2 The Proof of Results

Lemma 1. *Let Q be a rectangle with side length vector $\delta = (\delta_1, \ldots, \delta_d)$, $P_{\mathbf{r}} = \operatorname{span}\{x^{\mathbf{k}} : \mathbf{k} \in \mathbb{N}_0^d, k_j < r_j, 1 \le j \le d\}$. For each $f \in W_p^{\mathbf{r}}(Q)$, there exists a polynomial $g \in P_{\mathbf{r}}$ such that,*

$$\|f - g\|_{C(Q)} \le c \sum_{j=1}^{d} \delta_j^{r_j - \frac{d}{p}} \|\partial^{r_j \mathbf{e}_j} f\|_{L_p(Q)}. \tag{19}$$

where c is independent of δ and f.

Lemma 2. *Let $F \subseteq \mathcal{F}(\Omega, K)$ and $\tilde{F} \subseteq \mathcal{F}(\tilde{\Omega}, \tilde{K})$ be nonempty sets $\Gamma : F \to \tilde{F}$ be of the following form. For κ, $m^* \in \mathbb{N}$ there are mappings*

$$\begin{aligned} &\eta_j : \tilde{\Omega} \to \Omega, \\ &\beta \ : k \to \mathbb{Z}[0, 2^{m^*}), \\ &\varrho \ : \tilde{\Omega} \times \mathbb{Z}[0, 2^{m^*})^{\kappa} \to \tilde{K} \end{aligned} \tag{20}$$

such that for $f \in F$ and $s \in \tilde{\Omega}$

$$(\Gamma(f))(s) = \varrho(s, \beta \circ f \circ \eta_0(s), \ldots, \beta \circ f \circ \eta_{\kappa-1}(s)) \,. \tag{21}$$

If $\tilde{S} : \tilde{F} \to G$ is any mapping and $S = \tilde{S} \circ \Gamma$, then for each $n \in \mathbb{N}_0$

$$e_{2\kappa n}^q(S, F) \le e_n^q(\tilde{S}, \tilde{F}) \,. \tag{22}$$

Lemma 3. *Let Ω, K and $F \subseteq \mathcal{F}(\Omega, K)$ be nonempty sets, let $k \in \mathbb{N}_0$ and let $S_l : F \to \mathbb{R}$ $(l = 0, \ldots, k)$ be mappings. Define $S : F \to \mathbb{R}$ by $S(f) = \sum_{l=0}^{k} S_l(f)$ $(f \in F)$. Let $n_0, \ldots, n_k \in \mathbb{N}_0$.*

(i) *Assume $\theta_0, \ldots, \theta_k \geqslant 0$ and put $n = \sum_{l=0}^{k} n_l$. Then*

$$e_n^q(S, F, \sum_{l=0}^{k} \theta_l) \leq \sum_{l=0}^{k} e_{n_l}^q(S_l, F, \theta_l). \tag{23}$$

(ii) *Assume $v_0, \ldots, v_k \in \mathbb{N}$ satisfy $\sum_{l=0}^{k} e^{-v_l/8} \leq 1/4$. Put $n = \sum_{l=0}^{k} v_l n_l$. Then*

$$e_n^q(S, F) \leq \sum_{l=0}^{k} e_{n_l}^q(S_l, F). \tag{24}$$

Lemma 4. *Let S, $T : F \to G$ be any mappings. Then the following hold:*

(i) $e_n^q(T, F) \leq e_n^q(S, F) + \sup_{f \in F} |T(f) - S(f)|$.

(ii) *If $K = \mathbb{K}$ and S is a linear operator from $\mathcal{F}(D, K)$ to G, then for all $\lambda \in \mathbb{K}$*

$$e_n^q(S, \lambda F) = |\lambda| \cdot e_n^q(S, F)\,. \tag{25}$$

Then we recall the results about summation from [3, 5]. For $1 \leq p < \infty$, let L_p^N denote the Banach space of all functions $f : \mathbf{Z}[0, N) \to \mathbb{R}$, equipped with the norm

$$\|f\|_{L_p^N} = \left(\frac{1}{N} \sum_{i=0}^{N-1} |f(i)|^p \right)^{1/p}. \tag{26}$$

The summation operator $S_N : \mathcal{B}(L_p^N) \to \mathbb{R}$ is defined by

$$S_N f = \frac{1}{N} \sum_{i=0}^{N-1} f(i). \tag{27}$$

Theorem 3. *Let $1 \leq p < 2$, n, $N \in \mathbb{N}$ with $2 < n \leq c_1 N$,*

$$c \cdot n^{-1} \leq e_n^q(S_N, \mathcal{B}(L_2^N)) \leq c \cdot n^{-1} \log^{3/2} n \log\log n, \tag{28}$$

$$c \cdot \lambda_p(n, N) \leq e_n^q(S_N, \mathcal{B}(L_p^N)) \leq c \cdot \max(\log(n/\sqrt{N}), 1)^{2/p-1} \lambda_p(n, N) \tag{29}$$

where $\lambda_p(n, N) = \min(n^{-2(1-1/p)}, n^{-2/p} N^{2/p-1})$.

Proof of Theorem 1: Let n_0 be sufficiently large integer such that $n_0^{\frac{g(\mathbf{r})}{r_j}} > 2$, let $m_j(\mathbf{r}) = [n_0^{\frac{g(\mathbf{r})}{r_j}}]$, $j = 1, \ldots, d$. Define the real number P_0 as $P_0 = \sum_{j=1}^{d} \log m_j(\mathbf{r})$. After some simple calculations, we can yield

$$2^{P_0 g(\mathbf{r})} \leq c_1 m_j^{r_j} \leq c_2 2^{P_0 g(\mathbf{r})},\ j = 1, \ldots, d. \tag{30}$$

Now we subdivide D into $2^{P_0 l}$ congruent rectangles D_{li} with side length vector $(\frac{1}{m_1(\mathbf{r})}, \ldots, \frac{1}{m_d(\mathbf{r})})$. Let $\mathbf{s}_{li}$ denote the point in D_{li} with the smallest Euclidean norm. We introduce the following extension operator $E_{li} : \mathcal{F}(D, \mathbb{R}) \to \mathcal{F}(D, \mathbb{R})$ by

$$(E_{li} f)(\mathbf{s}) = f(\mathbf{s}_{li} + (m_1^{-l} s_1, \ldots, m_d^{-l} s_d)). \tag{31}$$

Let J be d-fold tensor product of appropriate degree Newton-Cotes formula with the form $Jf = \sum_{j=0}^{\kappa-1} b_j f(\mathbf{t}_j)$, where $\kappa = \dim P_{\mathbf{r}}$, which is exact on $P_{\mathbf{r}}$. By Lemma 1 and (30) that for $f \in W_p^{\mathbf{r}}(D)$

$$|I_d f - J_l f| \le c\, 2^{-P_0 g(\mathbf{r}) l} |f|_{W_p^{\mathbf{r}}(D)}. \tag{32}$$

Now we apply the multilevel techniques as in [4]. We approximate $I_d f$ by $J_k f$ with the desired error bound, however the node number required by J_k might be much larger than n. We split J_k into J_{k_0} and J'_l $(l = k_0, \ldots, k-1)$. The computation of $J'_l f$ reduces to that of the mean of sequences with bounded $L_p^{N_l}$ norms for proper N_l. This enable us to use Theorem 3 and approximate the mean by quantum algorithm.

Let $J'f := (J_1 - J_0)f$. Then $J'f$ has the form

$$J'f = \sum_{j=0}^{\kappa'-1} b'_j f(\mathbf{t'}_j) \tag{33}$$

where

$$\kappa' \le \kappa(2^{P_0} + 1)\,. \tag{34}$$

For $l \in \mathbb{N}_0$, set

$$J'_{li} f = J'(E_{li} f) \tag{35}$$

$$J'_l = 2^{-P_0 l} \sum_{i=0}^{2^{P_0 l}-1} J'_{li}\,. \tag{36}$$

We conclude that:

$$J_{l+1} f = 2^{-P_0 l} \sum_{i=0}^{2^{P_0 l}-1} J_1(E_{li} f) \tag{37}$$

and hence

$$J_{l+1} f - J_l f = 2^{-P_0 l} \sum_{i=0}^{2^{P_0 l}-1} J'_{li} f = J'_l f\,. \tag{38}$$

Using (35) and (40), we get

$$\begin{aligned} 2^{-P_0 l} \sum_{i=0}^{2^{P_0 l}-1} |J'_{li} f|^p &\le 2^{-P_0 l} \sum_{i=0}^{2^{P_0 l}-1} |J_1(E_{li} f) - J_0(E_{li} f)|^p \\ &\le 2^{-P_0 l} \sum_{i=0}^{2^{P_0 l}-1} (|(I_d - J_1)(E_{li} f)| + |(I_d - J_0)(E_{li} f)|)^p \\ &\le c\, 2^{-p g(\mathbf{r}) P_0 l} \|f\|^p_{W_p^{\mathbf{r}}(D)}\,. \end{aligned} \tag{39}$$

For $n \geq \max(\kappa, 5)$, let $k_0 = \lfloor \log(n/\kappa)/P_0 \rfloor$ and $k = \lceil (1 + 1/g(\mathbf{r}))k_0 \rceil$. By (2.40) we have

$$J_k = J_{k_0} + \sum_{l=k_0}^{k-1} J'_l. \tag{40}$$

For $k_0 \leq l < k$ put $N_l = 2^{P_0 l}$. To apply Lemma 3 we shall define the mapping $\Gamma_l : W_p^{\mathbf{r}}(D) \to L_p^{N_l}(D)$. Since $g(\mathbf{r}) > 1/p$, by the imbedding theorem [8]

$$\|f\|_{C(D)} \leq c\|f\|_{W_p^{\mathbf{r}}(D)}. \tag{41}$$

Let us fix an $m^* \in N$ with

$$2^{-m^*/2} \leq k^{-1} 2^{-g(\mathbf{r})kP_0} \tag{42}$$

and

$$2^{m^*/2-1} \geq c, \tag{43}$$

where the constant c comes from the inequality (41). Therefore for $f \in \mathcal{B}(W_p^{\mathbf{r}}(D))$

$$\|f\|_{C(D)} \leq 2^{m^*/2-1}. \tag{44}$$

Let the mapping $\eta_{l_j}(i) : \mathbb{Z}[0, N_l) \to D$ $(j = 1, \ldots, \kappa' - 1)$ be

$$\eta_{l_j}(i) = \mathbf{s}_{l_i} + (m_1^{-l} t'_{j1}, \ldots, m_d^{-l} t'_{jd}). \tag{45}$$

Define $\beta : R \to \mathbb{Z}[0, 2^{m*})$ by

$$\beta(z) := \begin{cases} 0 & \text{if } z < -2^{m^*/2-1}, \\ \lfloor 2^{m^*/2}(z + 2^{m^*/2-1}) \rfloor & \text{if } -2^{m^*/2-1} \leq z < 2^{m^*/2-1}, \\ 2^{m^*} - 1 & \text{if } z \geq 2^{m^*/2-1}, \end{cases} \tag{46}$$

and $\gamma : Z[0, 2^{m^*}) \to \mathbb{R}$ by

$$\gamma(y) = 2^{-m^*/2} y - 2^{m^*/2-1}. \tag{47}$$

It is obvious that for $-2^{m^*/2-1} \leq z \leq 2^{m^*/2-1}$,

$$\gamma(\beta(z)) \leq z \leq \gamma(\beta(z)) + 2^{-m^*/2}. \tag{48}$$

The mapping $\varrho : Z[0, 2^{m^*})^{\kappa'} \to \mathbb{R}$ is defined by

$$\varrho(y_0, \ldots, y_{\kappa'-1}) = \sum_{j=0}^{\kappa'-1} b'_j \gamma(y_j). \tag{49}$$

We give the expression of the mapping Γ_l for $f \in \mathcal{B}(W_p^{\mathbf{r}}(D))$,

$$\Gamma_l(f)(i) = \varrho((\beta \circ f \circ \eta_{l_j}(i))_{j=0}^{\kappa'-1}). \tag{50}$$

Let $\mathbf{x} = \mathbf{s}_{l_i} + (m_1^{-l} t'_{j1}, \ldots, m_d^{-l} t'_{jd})$ by (35) and (50) we have

$$|J'_{l_i} f - \Gamma_l(f)(i)| \le \sum_{j=0}^{\kappa'-1} |b'_j| |f(\mathbf{x}) - \gamma(\beta(f(\mathbf{x})))|. \tag{51}$$

By (44) and (48),

$$|J'_{l_i} f - \Gamma_l(f)(i)| \le 2^{-m^*/2} \sum_{j=0}^{\kappa'-1} |b'_j| \le ck^{-1} 2^{-g(\mathbf{r})kP_0}. \tag{52}$$

By (36)

$$|J'_l f - S_{N_l} \Gamma_l(f)| \le ck^{-1} 2^{-g(\mathbf{r})kP_0}. \tag{53}$$

Using (39), (53) and $l < k$, we have

$$\begin{aligned} \|\Gamma_l(f)\|_{L_p^{N_l}} &\le \|(J'_{l_i} f)_{i=0}^{N_l-1}\|_{L_p^{N_l}} + \|\Gamma_l(f) - (J'_{l_i} f)_{i=0}^{N_l-1}\|_{L_p^{N_l}} \\ &\le c \cdot 2^{-g(\mathbf{r})P_0 l}. \end{aligned} \tag{54}$$

We conclude that

$$\Gamma_l(\mathcal{B}(W_p^{\mathbf{r}}(D))) \subseteq c 2^{-g(\mathbf{r})P_0 l} \mathcal{B}(L_p^{N_l}). \tag{55}$$

Choose μ to satisfy the following condition

$$0 < \mu < \min\left(g(\mathbf{r})P_0, \frac{p}{2} P_0 \left(g(\mathbf{r}) - (\frac{2}{p} - 1) \right) \right) \tag{56}$$

and let

$$n_l = \lceil 2^{P_0 k_0 - \mu(l-k_0)} \rceil, \tag{57}$$

$$v_l = \lceil 8(2\ln(l - k_0 + 1) + \ln 8) \rceil, \tag{58}$$

for $l = k_0, \ldots, k-1$. A simple computation leads to

$$\sum_{l=k_0}^{k-1} e^{-v_l/8} < 1/4\,. \tag{59}$$

Set

$$\tilde{n} = n + 2\kappa' \sum_{l=k_0}^{k-1} v_l n_l. \tag{60}$$

By the definitions of n_l, v_l, k_0 and k,$\tilde{n} \le cn$. Using Lemma 4 (i) and (32), we have

$$e_{\tilde{n}}^q(I_d, \mathcal{B}(W_p^{\mathbf{r}}(D))) \le c 2^{-g(\mathbf{r})P_0 k} + e_{\tilde{n}}^q(J_k, \mathcal{B}(W_p^{\mathbf{r}}(D))). \tag{61}$$

Since $\kappa 2^{P_0 k_0} \le n$,

$$e_n^q(J_{k_0}, \mathcal{B}(W_p^{\mathbf{r}}(D)), 0) = 0. \tag{62}$$

According to the Lemma 3, (62) and Lemma 4, (44), (61), (62),

$$\begin{aligned} e_{\tilde{n}}^q(J_k, \mathcal{B}(W_p^{\mathbf{r}}(D))) &\le e_n^q(J_{k_0}, \mathcal{B}(W_p^{\mathbf{r}}(D))), 0) + e_{\tilde{n}-n}^q(J_k - J_{k_0}, \mathcal{B}(W_p^{\mathbf{r}}(D))) \\ &\le \sum_{l=k_0}^{k-1} e_{2\kappa' n_l}^q(J_l', \mathcal{B}(W_p^{\mathbf{r}}(D))). \end{aligned} \tag{63}$$

From (53), Lemma 4, (55), and Lemma 2, we have

$$\begin{aligned} e_{2\kappa' n_l}^q(J_l', \mathcal{B}(W_p^{\mathbf{r}}(D))) &\le ck^{-1}2^{-g(\mathbf{r})P_0 k} + e_{2\kappa' n_l}^q(S_{N_l}\Gamma_l, \mathcal{B}(W_p^{\mathbf{r}}(D))) \\ &\le ck^{-1}2^{-g(\mathbf{r})P_0 k} + c2^{-g(\mathbf{r})P_0 l} e_{n_l}^q(S_{N_l}, \mathcal{B}(L_p^{N_l})). \end{aligned} \tag{64}$$

Combining (61)-(64), we conclude

$$e_{\tilde{n}}^q(I_d, \mathcal{B}(W_p^{\mathbf{r}}(D))) \le c2^{-g(\mathbf{r})P_0 k} + c\sum_{l=k_0}^{k-1} 2^{-g(\mathbf{r})P_0 l} e_{n_l}^q(S_{N_l}, \mathcal{B}(L_p^{N_l})). \tag{65}$$

The treat of the case of $2 < p < \infty$ is similar to $p = \infty$, cf. [6]. We omit it. Next we consider the case $1 \le p < 2$, Note that by (56)

$$\frac{2}{p}\mu < P_0\left(g(\mathbf{r}) - (\frac{2}{p} - 1)\right). \tag{66}$$

For brief, denote $F = \mathcal{B}(W_p^{\mathbf{r}}(D))$. From (65) and Theorem 3

$$\begin{aligned} e_{\tilde{n}}^q(I_d, F) &\le c2^{-g(\mathbf{r})P_0 k} + c\sum_{l=k_0}^{k-1} 2^{-P_0 g(\mathbf{r})l - \mu(l-l_0) + (2/p-1)P_0 l}(k_0+1)^{\frac{2}{p}-1} \\ &\le c \cdot 2^{-(g(\mathbf{r})+1)P_0 k_0}(k_0+1)^{\frac{2}{p}-1} \\ &\le c \cdot n^{-g(\mathbf{r})-1}(\log n)^{\frac{2}{p}-1}. \end{aligned} \tag{67}$$

Finally we consider the case $p = 2$. From (65) and Theorem 3

$$\begin{aligned} e_{\tilde{n}}^q(I_d, F) &\le c2^{-g(\mathbf{r})P_0 k} + \sum_{l=k_0}^{k-1} c \cdot 2^{-g(\mathbf{r})P_0 l} \cdot n_l^{-1}\lambda_p(n_l, N_l)^{3/2}\log\lambda_p(n_l, N_l) \\ &\le c \cdot n^{-g(\mathbf{r})-1}\lambda_0(n). \end{aligned} \tag{68}$$

Combine the method in [4] and the skill of treating anisotropy, we prove the lower bound, so the proof of Theorem 1 is complete.

Based on Theorem 1, using the method in [2], we can prove Theorem 2.

References

[1] W. Dahmen, R. DeVore, K. Scherer, Multi-dimensional spline approximation. SIAM J. Numer. Anal., **17**, 380-402, 1980.

[2] Fang, Gensun and Ye, Peixin: Complexity of the deterministic and randomized methods for multivariate Integration problem for the Class $H_p^{\Lambda}(I^d)$, IMA. J. Numer. Anal. 2005 25:473-485.

[3] S. Heinrich, Quantum summation with an application to integration, J. Complexity **18**, 1-50, 2002.

[4] S. Heinrich, Quantum integration in Sobolev classes. J. Complexity **19**, 19-42, 2003.

[5] S. Heinrich, Erich Novak, On a problem in quantum summation, J. Complexity **18**, 1-18, 2003.

[6] Hu Xiaofei, Ye Peixin: Quantum Complexity of the Integration Problem for Anisotropic Classes, J. Comput. Math. 23, 3, 2005, 233-246.

[7] M.A. Nielsen, I.L. Chuang, Quantum Computation and Quantum Information. Cambridge University Press, Cambridge, 2000.

[8] S. M. Nikolskii, Approximation of functions of several variables and imbedding theorems, Springer-Verlag, 1975.

[9] E. Novak, Quantum complexity of integration. J. Complexity, **17**, 2-16, 2001.

[10] P.W. Shor, Introduction to Quantum Computing Algorithms, Birkhauser, Bonston, 1999.

[11] V.N. Temlyakov, Approximation of periodic functions, New York, Nova Science, 1993.

Quantum ANDOS Protocol with Unconditional Security*

Wei Yang, Liusheng Huang, Mingjun Xiao, and Weiwei Jing

Department of Computer Science and Technology
University of Science and Technology of China
National High Performance Computing Center (Hefei)
Hefei, 230027, P.R. China
smartyw@mail.ustc.edu.cn, lshuang@ustc.edu.cn,
xiaomj@ustc.edu.cn, wwjing@mail.ustc.edu.cn

Abstract. All-or-Nothing Disclosure of Secrets, or ANDOS for brevity, is an interesting cryptographic task, which involves two parties, say Alice and Bob. Alice has a few secrets, which she regards as equally valuable. She would like to sell any of them to Bob, with the guarantee that no information about the other secrets will be obtained if he pays for only one secret. Moreover, Bob can freely choose his secret and wants to ensure that Alice can obtain no information about which of her secrets he has picked. In this paper, we present a new quantum ANDOS scheme which achieves the same functionality, but which is of unconditional security and is able to contain arbitrary number of secrets.

1 Introduction

In modern cryptography, secure two-party protocols are of great value both in research and in practice. All-or-Nothing Disclosure of Secrets (ANDOS) is such a kind of cryptographic task that involves two parties, Alice and Bob. Alice has a few secrets, which she regards as equally valuable. She would like to sell any of them to Bob, with the guarantee that no information about the other secrets will be obtained if he pays for only one secret. Moreover, Bob can freely choose his secret and wants to ensure that Alice can obtain no information about which of her secrets he has picked.

The first ANDOS protocol was introduced in 1986 by Brassard, Crépeau and Robert in [1]. Indeed, the first case known as *One-out-of-t* String Oblivious Transfer for $t = 3$ was addressed and solved in [2]. After [1] and [2], several other schemes were proposed. Salomaa and Santean [3] designed an efficient ANDOS protocol that can distribute secrets to any number of people greater than one. Adrian Kent [4] gave a novel solution to this problem using the exchange of quantum information. In [5], Stern proposed a very efficient algorithm for real-life applications.

However, previous ANDOS protocols all have some drawbacks. In [2] the number of secrets is limited to no more than three. [3] is vulnerable to collusion among the participants. Under cryptographic assumptions, the security obtained by [1] and [5] can

* Supported by the NSF of China (Grant No. 60573171).
* Supported (in part) by the National Grand Fundamental Research 973 Program of China (Grant No. 2003CB317000).

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 20 – 30, 2006.

complement each other: the former is computationally secure for Alice and unconditionally secure for Bob, the latter provides unconditional security to Alice and computational security to Bob. In [4], Adrian Kent claims that his protocol is unconditionally secure. However, he used a bit commitment protocol as a building block in his scheme. In fact, unconditionally secure bit commitment is known to be impossible in both the classical and quantum worlds（[6]～[7]）. Thus the protocol given in [4] does not really have the property of unconditional security.

This paper gives a new ANDOS protocol by virtue of the elegant nature of quantum *Positive Operator-Valued Measure* or POVM measurements. Comparing to previous ANDOS protocols, the major contributions of this work are:

① Both Alice and Bob have no limitation on their computing power, so it is unconditionally secure for each participants. (Comparing to [4], it need not invoke a bit commitment scheme as a building block in our scheme.)

② According to *Heisengberg Uncertainty Principle* and *quantum no-cloning theorem*, no eavesdropper can escape being detected. Obviously, this is impossible in a classical environment.

③ The number of secrets can be an arbitrary positive integer greater than 2.

The paper is organized as follows. The k-OT protocol based on POVM construc- ted as a sub protocol in our ANDOS scheme is presented in section 2. The k-OT protocol ensures that Bob can reliably get the right information of each bit sent by Alice with probability about 0.3. In section 3, we use the k-OT protocol to construct our ANDOS scheme and then we introduce the intuition behind the scheme. We prove that the ANDOS protocol presented in this paper cannot be cheated by either party, except with arbitrarily small probability in section 4. Section 5 concludes the paper with some open questions.

2 k-OT Sub Protocol Based on POVM Measurements

First let us define two constants to be used later: $\mu = \cos\frac{\pi}{8}$, $\upsilon = \sin\frac{\pi}{8}$.

In order to distinguish two non-orthogonal states $|\varphi_1\rangle = \left(-|0\rangle + |1\rangle\right)/\sqrt{2}$ and $|\varphi_2\rangle = |1\rangle$ reliably with success probability of some constant, consider a POVM containing three elements,

$$E_1 = \alpha \cdot \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)\left(\frac{\langle 0| + \langle 1|}{\sqrt{2}}\right) = \frac{\alpha}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \tag{1}$$

$$E_2 = \beta \cdot |0\rangle\langle 0| = \beta\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{2}$$

$$E_3 = \gamma \cdot \left(-\upsilon|0\rangle + \mu|1\rangle\right)\left(-\upsilon\langle 0| + \mu\langle 1|\right) = \gamma\begin{pmatrix} \upsilon^2 & -\mu\upsilon \\ -\mu\upsilon & \mu^2 \end{pmatrix}. \tag{3}$$

where α, β and γ are constants to be later determined.

Because operators E_1, E_2 and E_3 constitute a POVM, they satisfy:

$$E_1 + E_2 + E_3 = I \ . \tag{4}$$

Now we can get that, for state $|\varphi_1\rangle$, if Bob performs the measurement described by the POVM $\{E_1, E_2, E_3\}$, the probabilities of outcomes of E_1 and E_2 are 0 and $\beta/2$ respectively. Similarly, for state $|\varphi_2\rangle$, the probabilities of outcomes of E_1 and E_2 are $\alpha/2$ and 0 respectively.

Therefore, suppose Alice sends to Bob a photon with the state $|\varphi_1\rangle$ or $|\varphi_2\rangle$, Bob measures it by the POVM $\{E_1, E_2, E_3\}$, if the result of his measurement is E_1 then Bob can safely conclude that the state he received must have been $|\varphi_2\rangle$. A similar line of reasoning shows that if the measurement outcome E_2 occurs then it must have been the state $|\varphi_1\rangle$ that Bob received. Some of the time, however, Bob will obtain the measurement outcome E_3, then he will infer nothing about the identity of the state he was given.

For the sake of symmetry of our protocol presented later, we let

$$\frac{\beta}{2} = \frac{\alpha}{2} \ . \tag{5}$$

Combine（1）～（5）we can get

$$\alpha = \beta = \frac{\sqrt{2}}{\sqrt{2}+1} \ . \tag{6}$$

$$\gamma = \frac{2}{\sqrt{2}+1} \ . \tag{7}$$

Let $k = \frac{\alpha}{2} = \frac{1}{2} \cdot \frac{\sqrt{2}}{\sqrt{2}+1} = \frac{1}{2+\sqrt{2}} \approx 0.292895$.

Now we can conclude that no matter what state Alice sends to Bob, he will always confirm the identity of the state he received with success probability of k.

Below we present the k-OT sub protocol.

Before the protocol, Alice and Bob agree on that $|\varphi_1\rangle = \left(-|0\rangle + |1\rangle\right)/\sqrt{2}$ and $|\varphi_2\rangle = |1\rangle$ represent the bit 0 and 1 respectively. $E_1 \sim E_3$ are defined as（1）~（3）, and α, β and γ are defined as（6）～（7）.

Protocol 1 k-OT_POVM (b)

Step 1. Let b denotes the bit Alice wants to send. If $b = 0$, then she sends Bob $|\varphi_1\rangle$. Otherwise she sends Bob $|\varphi_2\rangle$.

Step 2. After received the state, Bob performs the POVM measurements with $\{E_1, E_2, E_3\}$.

It is easy to know that protocol k-OT_POVM (b) satisfy:

① Alice knows $b=0$ or $b=1$.
② Bob gets bit b from Alice with probability k.
③ Alice does not know whether Bob got b rightly or not.

3 The New ANDOS Protocol

In this section we present a new quantum ANDOS scheme based on protocol 1 and introduce the intuition behind the scheme.

Before presenting our new scheme, we make the following assumption as in [1]: we assume that Alice is honest when she claims to be willing to disclose one secret, that is, she is not going to send junk to Bob.

3.1 Quantum ANDOS Protocol Based on POVM Measurements

Let t be the number of Alice's secrets, and $s_1, s_1, \cdots, s_t$ be the secrets themselves. Let c be Bob's choice (i.e. Bob wants to get s_c). In order for Alice to sell one secret to Bob, she uses protocol Quantum_ANDOS($s_1, s_1, \cdots, s_t$) (c) with Bob.

Before the protocol, Alice and Bob agree on a security parameter N used below.

Protocol 2 Quantum_ANDOS ($s_1, s_1, \cdots, s_t$)(c)

Step 1. Alice selects uniformly and randomly a N-bit sting $R = r_1 r_2 \cdots r_N$.

Step 2. For each bit in R, Alice uses the k-OT_POVM (b) protocol to transmit the relevant state to Bob.

Step 3. Bob records each state he measured. If $t \le 5$ then they skip to step 5 directly. Otherwise Bob sends ω bits of the measurement outcomes that he has measured with certainty to Alice, where $\omega = \left\lceil \frac{kt-1.2}{t-1.2} N \right\rceil$ will be explained later.

Step 4. Alice verifies that each bit she received be equal to the corresponding bit in R. If Bob pass the verification then they go to step 5. Otherwise she outputs "reject" to terminate the protocol.

Step 5. Bob partitions his measurement outcomes (which were not sent to Alice) into t sets $B_i = \left\{ a_{\theta(i-1)+1}, a_{\theta(i-1)+2}, \cdots a_{\theta i} \right\}$ $\left(i \in [1,t] \right)$, where θ is defined as follows:

$$\theta = \begin{cases} \lceil 0.19N \rceil & \text{if } t \le 5 \\ \dfrac{N-\omega}{t} & \text{otherwise} \end{cases}$$

where N is selected such that $\frac{N-\omega}{t}$ would be an integer. Bob is sure that he knows every r_{a_j} affirmatively for each $a_j \in B_c (j \in [\theta(c-1)+1, \theta c])$.

Step 6. Bob sends t-tuple $(X_1, X_2, \cdots, X_t) = (B_1, B_2, \cdots, B_t)$ to Alice.

Step 7. For each $i \in [1,t]$, Alice gets a binary string $m_i = r_{a_{\theta(i-1)+1}} r_{a_{\theta(i-1)+2}} \cdots r_{a_{\theta i}}$. Then she sends Bob t-tuple $(Y_1, Y_2, \cdots, Y_t) = (s_1 \oplus m_1, s_2 \oplus m_2, \cdots, s_t \oplus m_t)$.

Step 8. Bob computes $Y_c \oplus m_c$ to get his secret sting s_c.

3.2 Intuition Behind the Protocol

Intuitively, in order for Bob to get only one of Alice's secrets, his sets of B's should satisfy that he knows at most one of the sets completely. So if t is large to some extent then Bob has to announce some measurement outcomes to Alice to give her adequate confidence that his knowledge about R is limited to a small number of bits. Therefore the ratio of p to q should be greater than $1/t$ and less than $2/t$, where p represents the number of Bob's secret affirmative bits and q represents the number of all his bits (exclusive of the disclosed bits in step 3). In our scheme, we let this ratio be $1.2/t$. Thus ω can be calculated as follows:

$$\frac{kN-\omega}{N-\omega} = \frac{1.2}{t}$$

then we get that $\omega = \frac{kt-1.2}{t-1.2}N$. Because ω must be an integer, we let $\omega = \left\lceil \frac{kt-1.2}{t-1.2}N \right\rceil$.

N is chosen large enough as a security parameter. Moreover, N satisfies that $\frac{N-\omega}{t}$ be an integer in our quantum ANDOS scheme. In fact, we can let $\theta = \left\lfloor \frac{N-\omega}{t} \right\rfloor$ for arbitrary N when $t \geq 5$, but obviously the former will make the proof of protocol 2 more convenient than the latter.

4 Analysis and Proofs

At the end of the above ANDOS protocol, Bob's knowledge about Alice's secrets can be divided into three cases:

A_1 : Bob obtains nothing about Alice's secrets.

A_2 : Bob gets only one of Alice's secrets.

A_3 : Bob gets more than one secret.

Clearly, the above three mutually exclusive events constitute a complete event group. What is the probability that each event occurs? Indeed, we show that

Theorem 1. For sufficiently large N, there exist a constant λ ($0<\lambda<1$) such that Bob can obtain at least one secret with probability at least $1-\lambda^N$.

In order to prove Theorem 1 we need an inequality named "Chernoff Bound"[8]. We present the inequality first and then come back to the proof.

Lemma 2. Chernoff Bound [8]. Let $p \le 0.5$, and let $X_1, X_2, \cdots, X_n$ be independent $0-1$ random variables, so that $\Pr[X_i = 1] = p$ for each i. Then for all ε, $0 < \varepsilon \le p(1-p)$, we have

$$\Pr\left[\left|\frac{\sum_{i=1}^{n} X_i}{n} - p\right| > \varepsilon\right] < 2 \cdot e^{-\frac{\varepsilon^2}{2p(1-p)} \cdot n} . \tag{8}$$

Proof of Theorem 1. Let

$$x_i = \begin{cases} 1 & \text{if Bob got } r_i \text{ reliably} \\ 0 & \text{otherwise} \end{cases} \quad (i \in [1, N]) . \tag{9}$$

By definition $\Pr[x_i = 1] = k$, $\Pr[x_i = 0] = 1-k$, and $\sum_{i=1}^{N} x_i$ indexes the total number of the bits Bob reliably got from $R = r_1 r_2 \cdots r_N$.

If we let $\varepsilon = k - 0.2 \approx 0.092895$, then by inequality (8) we get

$$\Pr\left[\left|\frac{\sum_{i=1}^{N} x_i}{N} - k\right| > k - 0.2\right] < 2 \cdot e^{-\frac{(k-0.2)^2}{2k(1-k)} \cdot N} \approx 2e^{-0.020833N} . \tag{10}$$

Similarly, if we let $\varepsilon = 0.36 - k \approx 0.067105$, then we get

$$\Pr\left[\left|\frac{\sum_{i=1}^{N} x_i}{N} - k\right| > 0.36 - k\right] < 2 \cdot e^{-\frac{(0.392895-k)^2}{2k(1-k)} \cdot N} \approx 2e^{-0.010871N} . \tag{11}$$

For integer $t > 5$, if we let $\varepsilon = 0.2\dfrac{1-k}{t-1.2}$, then

$$\Pr\left[\left|\frac{\sum_{i=1}^{N} x_i}{N} - k\right| > 0.2\frac{1-k}{t-1.2}\right] < 2 \cdot e^{-\frac{\left(0.2\frac{1-k}{t-1.2}\right)^2}{2k(1-k)} \cdot N} \approx 2e^{-\frac{0.048284}{(t-1.2)^2} N} . \tag{12}$$

At last, if we let $\varepsilon = 0.8\dfrac{1-k}{t-1.2}$, then we get

$$\Pr\left[\left|\frac{\sum_{i=1}^{N} x_i}{N} - k\right| > 0.8\frac{1-k}{t-1.2}\right] < 2\cdot e^{-\frac{\left(0.8\frac{1-k}{t-1.2}\right)^2}{2k(1-k)}\cdot N} \approx 2e^{-\frac{0.772544}{(t-1.2)^2}N} . \quad (13)$$

If $t \le 5$, by inequality (10) we get

$$\Pr[Bob\ \ gets\ \ at\ \ least\ \ one\ \ secret] = 1-\Pr[A_1] = 1-\Pr\left[\sum_{i=1}^{N} x_i < \theta\right]$$

$$= 1-\Pr\left[-\frac{\sum_{i=1}^{N} x_i}{N} > -\frac{\theta}{N}\right] = 1-\Pr\left[k-\frac{\sum_{i=1}^{N} x_i}{N} > k-0.2+\frac{0.2N-\theta}{N}\right]$$

$$\ge 1-\Pr\left[k-\frac{\sum_{i=1}^{N} x_i}{N} > k-0.2\right] \ge 1-\Pr\left[\left|k-\frac{\sum_{i=1}^{N} x_i}{N}\right| > k-0.2\right] > 1-2e^{-0.020833N}$$

As long as N is large enough

$$1-2e^{-0.020833N} > 1-\left(e^{-0.020}\right)^N$$

Therefore for any constant λ, $e^{-0.020} < \lambda < 1$, Theorem 1 follows.

If $t > 5$, by inequality (12) we get that

$$\Pr[Bob\ \ gets\ \ at\ \ least\ \ one\ \ secret] = 1-\Pr[A_1] = 1-\Pr\left[\sum_{i=1}^{N} x_i - \omega < \theta\right]$$

$$= 1-\Pr\left[-\frac{\sum_{i=1}^{N} x_i}{N} > -\frac{\theta+\omega}{N}\right] = 1-\Pr\left[k-\frac{\sum_{i=1}^{N} x_i}{N} > k-\frac{\omega+\frac{N-\omega}{t}}{N}\right]$$

$$= 1-\Pr\left[k-\frac{\sum_{i=1}^{N} x_i}{N} > k-\frac{(t-1)\omega+N}{tN}\right] \ge 1-\Pr\left[k-\frac{\sum_{i=1}^{N} x_i}{N} > k-\frac{(t-1)\frac{kt-1.2}{t-1.2}N+N}{tN}\right]$$

$$= 1-\Pr\left[k-\frac{\sum_{i=1}^{N} x_i}{N} > 0.2\frac{1-k}{t-1.2}\right] \ge 1-\Pr\left[\left|k-\frac{\sum_{i=1}^{N} x_i}{N}\right| > 0.2\frac{1-k}{t-1.2}\right] \ge 1-2e^{-\frac{0.048284}{(t-1.2)^2}N}$$

As long as N is large enough

$$1-2e^{-\frac{0.048284}{(t-1.2)^2}N} > 1-\left(e^{-\frac{0.048}{(t-1.2)^2}}\right)^N$$

Therefore for any constant λ, $e^{-\frac{0.048}{(t-1.2)^2}} < \lambda < 1$, Theorem 1 follows. □

Theorem 1 tells us that Bob can get at least one of Alice's secrets with a probability that can be made arbitrarily close to 1. Then, suppose Bob is semi-honest, will he obtain more than one secret sent by Alice? Or in other word, what is the probability that event A_3 occurs? In fact, we have

Theorem 3. For sufficiently large N, there exist a constant η $(0<\eta<1)$ such that Bob can obtain more than one secret with probability at most η^N.

Proof. Let x_i be defined as (9), if $t \le 5$,

$$\Pr[Bob\ \ gets\ \ more\ \ than\ \ one\ \ secret] = \Pr[A_3] = \Pr\left[\sum_{i=1}^{N} x_i \ge 2\theta\right]$$

$$\le \Pr\left[\sum_{i=1}^{N} x_i > 2\cdot 0.18N\right] = \Pr\left[\frac{\sum_{i=1}^{N} x_i}{N} - k > 0.36 - k\right] \le \Pr\left[\left|\frac{\sum_{i=1}^{N} x_i}{N} - k\right| > 0.36 - k\right] < 2e^{-0.010871N}$$

where the last inequality uses inequality (11).

As long as N is large enough, the probability that Bob will get more than one secret from Alice can be made arbitrarily small.

Similar to the proof of Theorem 1, for any constant η, $e^{-0.010} < \eta < 1$, Theorem 3 follows.

Otherwise, if $t > 5$, by inequality (13) we get that

$$\Pr[Bob\ \ gets\ \ more\ \ than\ \ one\ \ secret] = \Pr[A_3] = \Pr\left[\sum_{i=1}^{N} x_i - \omega \ge 2\theta\right]$$

$$= \Pr\left[\sum_{i=1}^{N} x_i \ge 2\frac{N-\omega}{t} + \omega\right] = \Pr\left[\sum_{i=1}^{N} x_i \ge \frac{2N + (t-2)\omega}{t}\right] \le \Pr\left[\sum_{i=1}^{N} x_i \ge \frac{2N + (t-2)\frac{kt-1.2}{t-1.2}N}{t}\right]$$

$$= \Pr\left[\frac{\sum_{i=1}^{N} x_i}{N} - k \ge 0.8\frac{(1-k)}{t-1.2}\right] \le \Pr\left[\left|\frac{\sum_{i=1}^{N} x_i}{N} - k\right| \ge 0.8\frac{(1-k)}{t-1.2}\right] \le 2e^{-\frac{0.772544}{(t-1.2)^2}N}$$

Similarly, as long as N is large enough, the probability that Bob will get more than one secret from Alice can be made arbitrarily small. Therefore for any constant η, $e^{-\frac{0.772}{(t-1.2)^2}} < \eta < 1$, Theorem 3 follows. □

Theorem 1 and 3 ensure that after protocol 2, an honest or semi-honest Bob can get only s_c from Alice. Now let us consider for a cheating Bob, if he cheats by measuring Alice's states in bases other than $\{E_1, E_2, E_3\}$, will he successfully get more than one secret?

Theorem 4. Even if cheating Bob performs other measurements instead of POVM, it will help him little to get more than one secret.

Proof. Besides POVM measurement, Bob can perform projective measurements that may maximize his information about each bit in sting R.

Suppose the measurement basis Bob chooses is " + ", then for state $|\varphi_1\rangle = (-|0\rangle + |1\rangle)/\sqrt{2}$, the probabilities of outcomes of $|0\rangle$ and $|1\rangle$ are both $1/2$. Similarly, if the state Alice sent is $|\varphi_2\rangle = |1\rangle$, the probabilities of outcomes of $|0\rangle$ and $|1\rangle$ are 0 and 1 respectively. Therefore if *R is* uniformly distributed, Bob obtains each of Alice's states with probability

$$\frac{1}{2}\times\frac{1}{2}+\frac{1}{2}\times 1 = 0.75$$

At a first glance, the result seems better than the success probability *k* in POVM measurements. But notice that the probability 0.75 here does not convince Bob that he has got the state reliably. Indeed, there exist only one case that Bob believes he has obtained the state rightly, i.e. the result of Bob's measurement is $|0\rangle$. In this case, Bob can infer that the state Alice sent must have been $|\varphi_1\rangle$. However, the probability of this case is only $1/4$, which is even worse than k.

If Bob chooses "$\times$" basis to perform the measurements, the result is similar. We omit the analysis for the sake of brevity.

However, if Bob's choice is non-canonical bases, then for whether $|\varphi_1\rangle$ or $|\varphi_2\rangle$ all the measurement outcomes will larger than 0, which will render Bob unsure of any state. □

Below we shall show that there is very little Alice can do in order to cheat in protocol 2.

Theorem 5. Alice knows noting about Bob's choice c in protocol 2.

Proof. In fact, Bob does not reveal anything that involves *c* until Step 4. Moreover, B_c is purely random and information-theoretically hidden from Alice for that she is unable to distinguish which of Bob's set he had measured with affirmative outcomes. Therefore, sending t-tuple $(X_1, X_2, \cdots, X_t) = (B_1, B_2, \cdots, B_t)$ to Alice at step 6 does not reveal anything about *c* either. Thus it is information-theoretically impossible for Alice to cheat, regardless of her computing power and available technology.

On the other hand, even if Alice deviated protocol 2 by sending entangled states to Bob, it will not help her to tell which secret Bob has obtained. Suppose Alice sends $|\varphi'_2\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ instead of $|\varphi_2\rangle = |1\rangle$. Bob receives it and performs POVM measurement as usual. By equation (2) and (6) we know that, the measurement outcomes no longer have the property of infallibility. This will render Bob unable to decide a state correctly. The subsequence is that Bob will obtain none of Alice's secrets. By the assumption mentioned at the start of this section the whole of the scheme will fall to the ground. Obviously, this kind of attack is trivial and Alice will not take it. □

Finally, we would like to discuss the wire-tapping detecting ability of protocol 2. Suppose there exists an eavesdropper, say Eve, on the channel, who tries to eavesdrop on the transmission. For each state, Eve will not be able to "read" it without altering it. Each state sent by Alice is converted into electrical energy as it is measured and destroyed, so Eve must generate a new state to send to Bob. By the proof of Theorem 4 it is clear that Eve's best strategy is to perform the same POVM measurements adopted by Bob. For each state sent by Alice, the probability of Eve's failing to confirm its state is $1-k$. Then Eve has to guess a significant number of states randomly. When Bob receives the states sent by Eve, his probability of getting the right information of each state is equal to $\left(k+\frac{1-k}{2}\right)\cdot k \approx 0.189341$, which is a value much less than k. Therefore by comparing small quantities of their bits publicly, Alice and Bob can reach a conclusion. If they find more differences than can be attributed to known sources, they will know that there is an eavesdropper on the channel and they will terminate the ANDOS protocol.

5 Conclusion and Open Questions

We have described a complete protocol for ANDOS based on POVM measurements. We have shown that in the light of the laws of quantum mechanics, this protocol cannot be cheated by either party except with exponentially small probability. The protocol in this paper does not invoke any bit commitment protocol and any eavesdropper can be detected efficiently. Furthermore, both Alice and Bob have no limitation on their computing power in our scheme. Therefore our ANDOS protocol is unconditionally secure.

In order to make the analysis easier, we present our new quantum ANDOS scheme without the consideration of transmission errors on the channel. In fact, it is impractical to some extent. How to construct a secure ANDOS protocol based on POVM measurements that can tolerate transmission errors? We will leave it as an open question.

References

1. G. Brassard, C. Crépeau, and Jean-Marc Robert: All-or-nothing disclosure of secrets. In A.M. Odlyzko, editor, Proc. CRYPTO 86, pages 234-238.Springer-Verlag, (1987). Lecture Notes in Computer Science No. 263.
2. S. Wiesner.: Conjugate coding. In Sigact News, volume 18, pages 78-88,1983. Original manuscript written circa (1970)
3. A. Salomaa and L. Santean: Secret selling of secrets with several buyers. In 42th EATCS Bulletin, pages 178-186, (1990)
4. Adrian Kent: Unconditionally secure all-or-nothing disclosure of secrets. Arxiv preprint quant-ph/9711025, (1997)
5. Julien P. Stern: A new and efficient all-or-nothing disclosure of secrets protocol. ASIACRYPT 1998, pages 357-371. (1996)

6. Hoi-Kwong Lo, H.F.Chau: “Is Quantum Bit Commitment Really Possible?”. Los Alamos preprint archive quant-ph/9603004, (1996)
7. Dominic Mayers: Unconditionally Secure Quantum Bit Commitment is Impossible. Fourth Workshop on Physics and Computation-PhysComp ‘96, (1996)
8. Oded Goldreich: Foundation of Cryptography Basic Tools. Cambridge University Press, (2001)

A Novel Immune Clonal Algorithm

Yangyang Li[1] and Fang Liu[2]

[1] Institute of Intelligent Information Processing, Xidian University, Xi'an 710071, China
[2] Computer school, Xidian University, Xi'an 710071, China
lyy_791@yahoo.com.cn

Abstract. This paper proposes a novel immune clonal algorithm, called a quantum-inspired immune clonal algorithm (QICA), which is based on the concept and principles of quantum computing, such as a quantum bit and superposition of states. Like other evolutionary algorithms, QICA is also characterized by the representation of the antibody (individual), the evaluation function, and the population dynamics. However, in QICA, antibody is proliferated and divided into a set of subpopulation groups. Antibodies in a subpopulation group are represented by multi-state gene quantum bits. In the antibody's updating, the scabilitable quantum rotation gate strategy and dynamic adjusting angle mechanism are applied to guide searching. Theoretical analysis has proved that QICA converges to the global optimum. Some simulations are given to illustrate its efficiency and better performance than its counterpart.

1 Introduction

Immune clonal algorithm (ICA) [1] is principally a stochastic search and optimization method based on the clonal selection principle in AIS. Compared to traditional optimization methods, such as calculus-based and enumerative strategies, ICA are robust, global, and may be applied generally without recourse to domain-specific heuristics. In particular, ICA has the better ability of the local search. But with the scale of the problem increased, ICA did not solve effectively the complicated problem.

This paper proposes a novel immune clonal algorithm, called a quantum-inspired immune clonal algorithm (QICA), which is based on merging quantum computing and clonal selection theory. Unlike other research areas, there has been relatively little work done in applying quantum computing to AIS. In [2], quantum-inspired computing was proposed. A quantum-inspired immune clonal algorithm was firstly introduced in [3] for solving the high dimensional function optimization problems. It should be noted that although QICA is based on the concept of quantum computing, QICA is not a quantum algorithm, but a novel optimization algorithm for a classical computer. There are two innovation points listed as follows in this paper.

1) Antibodies are proliferated and divided into some subpopulation. Antibodies in a subpopulation are represented by multi-state gene quantum bits. The quantum bit antibody has the advantage that it can represent a linear superposition of states (classical solutions) in search space probabilistically. Thus, the quantum bit representation has a better characteristic of population diversity than other representations.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 31 – 40, 2006.

2) In the antibody's updating, the scabilitable quantum rotation gate strategy and dynamic adjusting angle mechanism are applied to guide searching. Quantum mutation focuses on a wise guidance by the current best antibody. Thus, it can accelerate convergence.

2 QICA

2.1 Representation

In AIS, a number of different representations can be used to encode the solutions onto antibodies. The representation can be generally classified as: binary, numeric, and symbolic. QICA uses a new representation, a qubit for the probabilistic representation that is based on the concept of qubits and a qubit antibody as a string of qubits, which are defined below.

Definition 1: The probability amplitude of one qubit is defined with a pair of numbers (α, β) as

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \tag{1}$$

where $|\alpha|^2 + |\beta|^2 = 1$.

Definition 2: The phase of a qubit is defined with an angle ω as

$$\omega = \arctan(\beta / \alpha) \tag{2}$$

and the product $\alpha * \beta$ is represented with the symbol *d*, i.e.

$$d = \alpha * \beta, \tag{3}$$

where *d* stands for the quadrant of qubit phase ω. If *d* is positive, the phase ω lies in the first or third quadrant, otherwise, the phase ω lies in the second or fourth quadrant.

Definition 3: A qubit antibody as a string of *m* qubits is defined as:

$$\begin{pmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_m \\ \beta_1 & \beta_2 & \dots & \beta_m \end{pmatrix}, \tag{4}$$

where $|\alpha_l|^2 + |\beta_l|^2 = 1, (l = 1, 2, ..., m)$. So the phase of the *l*-th qubit is

$$\omega_l = \arctan(\beta_l / \alpha_l). \tag{5}$$

2.2 Immune Clonal Algorithm

The Clonal Selection Theory is put forward by Burnet in 1958 [4]. Based on the clonal selection theory, Immune Clonal Algorithm (ICA) is proposed. The used ICA in this paper is an antibody random map induced by the avidity including three steps:

clone operation, immune genetic operation and clonal selection operation. The state transfer of antibody population is denoted as follows [5]:

$$A(t) \xrightarrow{\text{clon operator}} A'(t) \xrightarrow{\text{immune genetic operator}} A''(t) \xrightarrow{\text{selection operator}} A(t+1).$$

2.3 The Quantum-Inspired Immune Clonal Algorithm (QICA)

Fig. 1 lists the major steps in the proposed algorithm.

Algorithm1: The quantum-inspired immune clonal algorithm
Step1: Initialize $Q(t)$ and $B(t)$, t=0.
Step2: Generate $Q'(t)$ from $Q(t)$ by the clonal operator Θ.
Step3: Update $Q'(t)$ by quantum mutation.
Step4: Produce $P'(t)$ by observing the updated $Q'(t)$.
Step5: Evaluate the avidity of $P'(t)$, produce $B(t)$ by clonal selecting $P'(t)$ and record the corresponding qubit to generate next generation $Q(t+1)$.
Step6: Store the best solutions among $B(t)$ to b and judge the termination condition
Step7: If it is satisfied, output the best solution and end the process; otherwise go to *step2*.

Fig. 1. The quantum-inspired immune clonal algorithm

The major elements of QICA are presented as follows.

- **The quantum population**

QICA maintains a quantum population $Q(t) = \{q_1^t, q_2^t, \cdots q_n^t\}$ at the t-th generation where n is the size of population, and m is the length of the qubit antibody q_i^t which is defined as: $q_i^t = \begin{Bmatrix} \alpha^t{}_1 & \alpha^t{}_2 & \ldots & \alpha^t{}_m \\ \beta^t{}_1 & \beta^t{}_2 & \ldots & \beta^t{}_m \end{Bmatrix}, i = 1, 2, ..., n$. At *step1* all α_l^t an β_l^t of q_i^t ($l = 1, 2, ..., m;\ t = 0$) are randomly generated between -1 and 1 satisfying $\left|\alpha_l^t\right|^2 + \left|\beta_l^t\right|^2 = 1, (l = 1, 2, ..., m).$

- **Clonal operator**

The clonal operator Θ is defined as:

$$\Theta(Q(t)) = [\Theta(q_1) \quad \Theta(q_2) \quad \cdots \quad \Theta(q_n)]^T, \tag{6}$$

where $\Theta(q_i) = I_i \times q_i, i = 1, 2 \cdots n$, and I_i is C_i dimension row vectors. Generally, C_i is given by:

$$C_i = g(N_c, D(q_i)) \qquad i = 1, 2 \cdots n, \tag{7}$$

and can be adjusted self-adaptively by the avidity D(*). N_c is a given value relating to the clone scale. After clone, the population becomes:

$$Q'(t) = \{Q(t), q'_{1,} \cdots q'_{n}\}, \tag{8}$$

where:

$$q'_i(t) = \{q_{i1}(t), q_{i2}(t), \cdots, q_{iC_i-1}(t)\}, q_{ij}(t) = q_i(t) \;\; j = 1, 2, \cdots, C_i - 1. \tag{9}$$

- **Immune gene operator**

Antibodies in population $Q'(t)$ are updated by applying quantum mutation defined as the immune genetic operator. The quantum mutation deduces a probability distribution in terms of the current best antibody in subpopulation, whose process is: define a guide quantum antibody $Q_{guide}(t)$ from the quantum rotation gate based on the current best antibody in subpopulation which is stored in $B(t)$ and spread the new antibody with this guide antibody being the center in subpopulation. Firstly, the quantum rotation gate is given as follows. Firstly, the quantum rotation gate is given as follows.

A quantum rotation gate U is

$$U(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}, \tag{10}$$

where θ is the rotate angle which controls the convergence speed and θ is defined as

$$\theta = k * f(\alpha_l, \beta_l), \tag{11}$$

where k is a coefficient determining the speed of convergence. If k is too big, search grid of the algorithm is large and the solutions may diverge or have a premature convergence to a local optimum; if it is too small on the other hand, search grid of the algorithm is also small and the algorithm may stay at a stagnant state. Here, by taking advantage of rapid convergence of QICA, k is defined as a variable that is relative to the clone scale. Thus, the better antibody quality is, the smaller the mutating size is and thus local search is the more advantageous. For example $k = 10 * \exp(-C_i / N_c)$, where C_i is the clone scale and N_c is a constant which is the clonal size.

The function $f(\alpha_l, \beta_l)$ determines the search direction of convergence to a global optimum. The following lookup table (Table 1) can be used as a strategy for convergence.

Table 1. Lookup Table of $f(\alpha_l, \beta_l)$

$d_1 > 0$	$d_2 > 0$	$f(\alpha_l, \beta_l)$	
		$\lvert\omega_1\rvert > \lvert\omega_2\rvert$	$\lvert\omega_1\rvert < \lvert\omega_2\rvert$
true	true	+1	-1
true	false	-1	+1
false	true	-1	+1
false	false	+1	-1

Table 1: $d_1 = \alpha_1 * \beta_1, \omega_1 = \arctan(\beta_1 / \alpha_1)$, *where* α_1, β_1 *is the probability amplitude of the best solution in subpopulation, and* $d_2 = \alpha_2^{'} * \beta_2^{'}$, $\omega_2 = \arctan(\beta_2^{'} / \alpha_2^{'})$, *where* $\alpha_2^{'}$, $\beta_2^{'}$ *is the probability amplitude of the current solution after clone.*

The quantum mutation can be described as

$$Q_{guide}(t) = U(\theta) * Q'(t), \tag{12}$$

$$Q''(t) = Q_{guide}(t) + I_i^{'} * a * N(0,1), \tag{13}$$

in which $I_i^{'}$ is $(C_i - 1)$ dimension unit row vectors. In other words, the quantum mutation is unused to $Q(t) \in Q'(t)$. a is the spread variance which often we let $a \in [0.05, 0.15]$ and $N(0,1)$ is a random number, chosen from a normal distribution with mean zero, variance one and standard deviation one.

- **Observing operator**

At *Step4*, in the act of observing a quantum state, it collapses to a single state (namely classical representation). For non-binary coding optimization problem, we observe the updated $Q'(t)$ (namely $Q''(t)$) and produce numerical stings population $P'(t) = \{x_1^t, x_2^t, \cdots, x_n^t\}$ where x_i^t $(i = 1, 2, ..., n)$ are numeric strings of length *m* derived from $\alpha_l^{''t}$ or $\beta_l^{''t}$ (*l*=1,..,*m*). The process is described as: (a) a random number $p \in [0,1]$; (b) if it is larger than $\left|\alpha_l^{''t}\right|^2$, the corresponding bit in $P'(t)$ takes $floor(\left|\beta_l^{''t}\right| \times 10)$; otherwise it takes $floor(\left|\alpha_l^{''t}\right| \times 10)$, where *floor*(*x*) rounds the elements of *x* to the nearest integer smaller than or equal to *x*. In fact, each bit in a antibody x_i^t denote a encoded variable and only the corresponding bit in x_i^t is mapped to the feasible solution space for converting from numeric stings to real value, namely a quantum antibody of length *m* is equal to the number of variable.

- **Clonal selection operator**

Assuming we will search maximal value of object function, the operation is given as follows: (a) for $\forall i = 1,2,\cdots n$, if there is the mutated and observed classical antibody b_i and $D(b_i) = \max\{D(x_{ij}) \mid j = 2,3,\cdots C_i\}$, namely: $D(b_i) > D(x'_i)$, then b_i replaces the antibody x'_i in the aboriginal population; (b) record the corresponding qubit of b_i as the next generation population $Q(t+1)$ at the same time. The antibody population is updated, and the information exchanging among the antibody population is realized.

3 Convergence of QICA

Definition 4: Assume that the size of a population is *n*, and *X* is a searching space which all the antibodies belong to. Let $X_t = (x_1, x_2, \cdots, x_n)$ be the population at the *t*-th generation, *D* be the avidity function (namely the fitness function) on *X* and f^* be the global optimum. Let:

$$M = \{\bar{X} \,|\, D(\bar{X}) = \max_{x_i \in X_t}\{D(x_i), i \le n\}\}, \tag{14}$$

M is called the satisfied set of population X_t.

Definition 5: Let $D_t = \max_{x_i \in X_t}\{D(x_i) : i = 1, 2, ..., n\}$. For any initial distribution, if the following equation holds,

$$\lim_{t \to \infty} P\{D_t = f^*\} = 1, \tag{15}$$

where *P* stands for the probability, then we say that the algorithm is convergent with probability 1.

Theorem 1: the population series of QICA $\{X_t, t \ge 0\}$ is a finite homogeneous Markov chain.

Proof: Like the evolutionary algorithms, the state transfer of QICA is processed on the finite space. Therefore, population is finite, since

$$X_{t+1} = T(X_t) = T_s \circ T_g \circ \Theta(X_t). \tag{16}$$

T_s, T_g and Θ denote the clonal selection operator, the immune genetic operator and the clone operator respectively. Note that T_s, T_g and Θ have no relation with *t*, so X_{t+1} is only related to X_t. Therefore, $\{X_t, t \ge 0\}$ is a finite homogeneous Markov chain.

Theorem 2: The M of Markov chain of QICA is monotonically increasing, namely, $\forall t \ge 0, D(X_{t+1}) \ge D(X_t)$.

Proof: It is apparent that the antibody of QICA does not degenerate since holding best strategy is used in the algorithm.

Theorem 3: The quantum-inspired immune clonal algorithm is convergent.

Proof: For Theorem 1 and Theorem 2, the QICA is convergent with the probability 1.

4 Numerical Optimization

4.1 Problem Definition

A global unconstrained numerical optimization can be formulated as solving the following objective function

$$\text{minimize} \quad f(x), \quad x = (x_1, ..., x_m) \in S, \tag{17}$$

where $S \subseteq R^m$ defines the search space which is an *m*-dimensional space bounded by the parametric constraints $\underline{x}_i \le x_i \le \bar{x}_i$, *i*=1,…*m*, *f*(*x*) is an objective function.

A global constrained numerical optimization can be formulated as solving the following objective function

$$\text{minimize} \quad f(x), \quad x = (x_1, ..., x_m) \in S \cap F \ , \tag{18}$$

where $S \subseteq R^m$ defines the search space which is an *m*-dimensional space bounded by the parametric constraints $\underline{x}_i \le x_i \le \overline{x}_i$, i=1,…m, $f(x)$ is an objective function and the *feasible region* F is defined by

$$F = \{x \in R^m \mid g_j(x) \le 0 \quad \forall j \in \{1, ..., N\} \} , \tag{19}$$

where $g_j(x), \ \forall j \in \{1, ..., N\}$ are *constraints*. In this paper, we transform a constrained optimization problem into an unconstrained one by the penalty term, such as the one given by (20):

$$\psi(x) = f(x) + A \sum_{j=1}^{N} \max\{0, g_j(x)\} \tag{20}$$

where A is static penalty coefficient. An antibody for numerical optimization problems is defined as follows:

Definition 6: An antibody, a, represents a candidate solution to the optimization problem in hand. The value of its avidity is equal to the negative value of the objective function,

$$a \in S \quad \text{and} \quad D(a) = -f(a). \tag{21}$$

4.2 Unconstrained Optimization

We execute the QICA to solve the unconstrained test functions f_1-f_4 (see [3]). f_1-f_4 are high-dimensional problems. They appear to be the most difficult class of problems for many optimization algorithms (including EP).

To identify any improvement due to qubit antibody design and quantum mutation operator, we design and carry out the following control experiment. We execute a standard immune clonal algorithm (SICA) to solve the above test functions. We adopt the following parameter values or scheme.

1) For QICA, the size of initial antibody population is 10 and the clonal sizes N_c is 30. The numbers of the qubits for the ten test functions are m, where m is dimensions of the above test functions and mutation probability p_m=0.9. The termination criterion is that the quality of the solution cannot be further improved in successive 50 generations for each function.
2) SICA (see section 2.2) does not apply qubit antibody design and quantum mutation operator. Gaussian mutation is used on the classical numeric representation. Its parameters are the same as those of QICA and the termination criterion is the same as that of QICA.

Table 2 shows the results averaged over 50 control trials. Recall that SICA is the same as QICA, except that it uses classical representation (numerical stings) and Gaussian mutation instead of qubit antibody and quantum mutation. We see that

SICA requires more function evaluations than QICA, and hence it has a larger time cost. However, SICA gives larger mean function values than QICA under the same termination criterion, and hence its mean solution quality is poorer. In addition, SICA gives larger standard deviations of function values than QICA, and hence its solution quality is less stable. These results indicate that quantum representation design and quantum mutation operator can effectively improve SICA.

Table 2. The comparing results of QICA and SICA on the function f_1-f_4, where m is the dimensions

f	m	Evaluations		Mean (St. Dev.)	
		QICA	SICA	**QICA**	SICA
f_1	30	**12,329**	108,431	**0 (0)**	5.841×10 (2.451×10^{-1})
f_2	100	**24,267**	343,710	**-78.3323 (2.339×10^{-10})**	-69.3441 (1.651×10^{-2})
f_3	30	**13,731**	225,213	**-12569.49 (1.035×10^{-10})**	-12554.5 (20.9)
f_4	100	**24,934**	356,341	**-99.5034 2.23×10^{-2}**	-86.8901 (2.626×10^{-1})

4.3 Constrained Optimization

In order to test the performance of QICA further, six constrained benchmark functions have been used (see [6]). f_7 and f_{10} included the equality constraints and All equality constraints have been converted into inequality constraints, $|h(x)-\delta|\le 0$, using the degree of violation $\delta = 10^4$. We transform a constrained optimization problem into an unconstrained one by the equation (20), where A is given by (22):

$$\begin{pmatrix} A_{f_5}=100 & A_{f_6}=10^4 & A_{f_7}=10 \\ A_{f_8}=500 & A_{f_9}=5\times10^6 & A_{f_{10}}=0.05 \end{pmatrix}. \tag{22}$$

For QICA, the size of initial antibody population is 20, the clonal sizes N_c is 60 and mutation probability p_m=0.9. The numbers of the qubits for the ten test functions are m, where m is dimensions of the above test functions.

RY [6] is a method proposed recently and obtains good performances on constrained optimization problems. In [6], the termination criterion of RY was to run 1750 generations for the six functions. Thus, to make a fair comparison, we compare both the qualities of their final solutions and the computational cost at the same. Accordingly, the termination criterion of QICA is that the quality of the solution cannot be further improved in successive 50 generations for each function. The results averaged over 100 trials for QICA are shown in Table 3, where the all solution for QICA is the feasible solution.

In Table 3, we compare the performance of QICA with RY, where including: 1) the best solution in the 100 trials (Best), 2) the mean function value (Mean) (i.e., the mean of the function values found in the 100 trials), 3) the standard deviation of function value (St. dev.), 4) the worst solution in the 100 trials (Worst), 5) the running time (Time) and 6) the mean number of function evaluations (Evaluations). Note that the results averaged over 30 trials for RY are shown in Table 3.

Table 3. Comparison between QICA and RY where the results for RY are obtained from [6]

f	Algorithm	Best	Mean	St. dev.	worst	Time (s)	Evaluations
f_5	**QICA**	**-0.803618**	**-0.787021**	**1.101×10^{-2}**	**-0.744321**	**1.85**	**64,904**
	RY	-0.803515	-0.781975	2.0×10^{-2}	0	-	350,000
f_6	**QICA**	**-30665.539**	**-30665.539**	**2.235×10^{-11}**	**-30665.539**	**0.43**	**23,112**
	RY	-30665.539	-30665.539	2.0×10^{-5}	-30665.539	-	350,000
f_7	**QICA**	**5126,497**	**5127,081**	**7.907×10^{-1}**	**5138,612**	**0.47**	**25,067**
	RY	5126,497	5128,881	3.5	5142,427	-	350,000
f_8	**QICA**	**680.630**	**680.631**	**1.231×10^{-3}**	**680.639**	**0.67**	**64,612**
	RY	680.630	680.656	3.4×10^{-2}	680.763	-	350,000
f_9	**QICA**	**7053.004**	**7183.128**	**54.1**	**7213.791**	**0.64**	**64,599**
	RY	7053.316	7559.613	530	8835.655	-	350,000
f_{10}	**QICA**	**0.053942**	**0.053964**	**2.015×10^{-7}**	**0.054002**	**0.5**	**26,109**
	RY	0.053957	0.067543	3.1×10^{-2}	0.216915	-	350,000

Tables 3 compare QICA with RY. As can be seen, QICA and RY can find the exact global optimum, in all trials for f_6, but QICA gives smaller standard deviation of function values than RY, and hence it has a more stable solution quality. For f_5, f_8, f_9 and f_{10}, both the solution of QICA are much better than those of RY. Moreover, the mean number of function evaluations of QICA is about 20 000 or 60 000 for all functions, while that of RY is 350 000. Therefore, the computational cost of QICA is much lower than those of OGA/Q. Especially, for all the test functions, the mean running time is between 1s and 2s. For f_7, the best solution of QICA is as good as that of RY, and both are smaller than the global optimum because equality constraints are converted into inequality constraints, but the mean and worst solution of QICA are better than those of RY. To summarize, the results show that QICA outperforms RY, and is competent for the numerical optimization problems.

5 Conclusions

This paper proposed a novel immune clonal algorithm-QICA, inspired by the concept of quantum computing. Our objective was to apply the quantum theory to enhance the immune clonal algorithm, so that it could be more robust and statistically sound. In particular, a qubit antibody was defined as a string of qubits for the probabilistic representation. The qubit representation has a better characteristic of population diversity

than other representations. Due to the novel representation, we put forward the quantum mutation operator which is associated with the probability amplitudes of basic quantum states and is used at the inner subpopulation to accelerate the convergence. Simulations are given to test this new algorithm, and rapid convergence and good global search capacity characterize the performance of QICA. The application of QICA to other problems such as the multiobjective optimization problem deserves our further research.

References

1. De Castro, L. N., Von Zuben, F. J.: Artificial Immune Systems: Part II—A Survey of Applications. FEEC/Univ. Campinas, Campinas, Brazil. [Online]. Available: http://www.dca.fee.unicamp.br/~lnunes/immune.html (2000)
2. Moore, M., Narayanan, A.: Quantum-Inspired Computing. Dept. Comput. Sci., Univ. Exeter, Exeter, U.K., (1995)
3. Li, Y.Y., Jiao, L.C.: Quantum-Inspired Immune Clonal Algorithm. in Proceedings of the 4th International Conference on Artificial Immune Systems, Christian Jacob, Marcin L. Pilat, Peter J. Bentley, et al, Eds. Banff, Alberta, Canada, Aug (2005) 304 – 317
4. Burnet, F. M.: Clonal Selection and After .In Theoretical Immunology, Bell, G. I., Perelson, A. S., pimbley Jr, g. H.(eds.) Marcel Dekker Inc., (1978) 63-85
5. Du, H.F., Jiao, L.C., Wang, S.A.: Clonal Operator and Antibody Clone Algorithms. in Proceedings of the First International Conference on Machine Learning and Cybernetics. Shichao, Z., Qiang, Y. and Chengqi, Z., Eds. IEEE, Beijing, (2002)506–510
6. Runarsson, T. P., Yao, X.: Stochastic Ranking for Constrained Evolutionary Optimization. IEEE Trans. Evol. Comput., Vol. 4, No. 3, Sep. (2000) 284-294

Secure Broadcasting Using the Secure Quantum Lock in Noisy Environments

Ying Guo[1,2], Guihua Zeng[1], and Yun Mao[2,3]

[1] The State Key Laboratory on Fiber-Optic Local Area Communication Networks and Advanced Optical Communication system, Department of Electronic Engineering, Shanghai Jiaotong University, Shanghai 200030, China
{yingguo1001, ghzeng}@sjtu.edu.cn
[2] Department of Communication Engineering, Central South University, Changsha 410200, China
[3] School of Science, Laiyang agriculture college, Qingdao 266109, China

Abstract. A secure quantum broadcasting scheme is proposed by means of a secure quantum lock, which is implemented by the legal communicators using the the generators of the stabilizer of a stabilizer quantum code (SQC). The advantages of the proposed scheme for the designed lock is that it can be implemented in the noisily broadcasting channel. Based on physics of the quantum system, the secret messages are broadcasted to the legal users in the noisy environments with unconditionally security.

1 Introduction

Different from Quantum key distribution (QKD), whose object is to establish a common key between the communicators, quantum broadcasting is to transmit the secret messages from a sender to several spatially separated receivers via quantum broadcasting channels [1]. Dubbed quantum telecloning has been suggested to distribute an input qubit into M particles, all spatially separated from each other. This transmission is achieved by first establishing a particular initial entangled state between the sender and the receiversto allow multiple receiver receivers at arbitrary location [2]. However, the structure of the distribution channels in these scheme are determined by the transmitted quantum messages and the illegal users may obtain the secret messages from the sender easily. Moreover, the secret messages can not be distributed in the noisy channel. To defend the illegal users against obtaining the messages from the sender and recover the destroyed qubits of the broadcasted state, a secure quantum lock is designed for the distribution the messages to several receivers in this paper. Except the legal users, none should unlock the broadcasted state.

Because the no-cloning theorem [3,4] forbids perfect duplication of quantum information by the receivers, it is possible for the sender to perform such one-to-many quantum communication perfectly with unconditional security by producing and sending several same quantum states at the same designed machine.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 41–44, 2006.

2 Description of the Quantum Broadcasting Scheme

Consider a broadcasting system consisting of a broadcast network and a group U containing m users. Let U_0 denote a nonempty subgroup of m_0 legal users within U for $2 \leq m_0 \leq m$. In order to send the k-qubit messages $|P\rangle = |p_1 \cdots p_k\rangle$ to U_0 with security, $|P\rangle$ should be enciphered before being broadcasted, in which a lock is superimposed onto the front of the messages to control the deciphering session keys that match the keys of U_0.

2.1 Construction of the Secure Lock

In this phase, we present the construction method of the secure lock L to generate the enciphering or deciphering session key K_s, K_v and K_{U_0} by exploiting a SQC, where K_{U_0} is private but K_v and K_s are public.

In the n-qubit depolarizing channel, error operators can be represented by Pauli group $G_n = \{\mu_i^{\otimes n}\}$, where $\mu_i \in \{I, X = \sigma_x, Z = \sigma_z, Y = ZX\}$ and $\otimes n$ denotes the n-fold tensor product [4]. SQC $[[n, k, d]]$ can be constructed from a stabilizer $\mathcal{S} = \{\prod_{j=1}^{n-k} M_j^{b_j} : b_j \in \{0, 1\}, 1 \leq j \leq n-k\}$, where $M_1, \cdots, M_{n-k} \in G_n$ are the (mutually commuting) generators of $\mathcal{S}$. Code words of SQC are generated by the eigenvectors associated with eigenvalue '1' of $\mathcal{S}$ denoted by$C(\mathcal{S}) = \{|\upsilon\rangle : M|\upsilon\rangle = |\upsilon\rangle, M \in \mathcal{S}\}$. Operators $\mathcal{Z}_1, \cdots, \mathcal{Z}_k$ can be selected from $\{\bar{Z} : \bar{Z}M = M\bar{Z}, M \in \mathcal{S}\}$ such that $\{M_1, \cdots M_{n-k}, \mathcal{Z}_1, \cdots \mathcal{Z}_k\}$ is an independently commuting set. For an arbitrary $\mathcal{Z}_i \in \Omega$, there is an operator $\mathcal{X}_i \in G_n$ satisfying $\mathcal{X}_i\mathcal{Z}_i = -\mathcal{X}_i\mathcal{Z}_i$ and $\mathcal{X}_i\sigma = \sigma\mathcal{X}_i$, where $\sigma \in \Omega$ for $\sigma \neq \mathcal{Z}_i$. Since the encoding on n-qubits can be written as tensor product of (single) qubit states, the encoder of the stabilized quantum code generates a k-qubit logical state [5]

$$|P\rangle_L = \frac{1}{\sqrt{2^{n-k}}}(\mathcal{X}_1^{p_1} \cdots \mathcal{X}_k^{p_k} \prod_{i=1}^{n-k}(I + M_i))|0^{\otimes n}\rangle). \tag{1}$$

Thus, the lock L, along with K_s, K_v and K_{U_0}, are generated as follows.

Step 1. Alice generates k operators $K_s = \{\mathcal{X}_1, \cdots, \mathcal{X}_k\}$, which are shared by the legal communicators.

Step 2. Alice regards $n-k$ generators of the stabilizer $\mathcal{S}$ of SQC as the lock L denoted by $L = \{M_1, M_2, \cdots, M_{n-k}\}$. She regards operators $K_v = \{M_{\tau+1}, \cdots, M_{n-k}\}$ as the public key used for deciphering the secret messages. Then, Alice sends the operator set $K_{U_0} = \{M_1', \cdots, M_\tau', M_{\tau+1}', \cdots M_{\tau+\tau_{U_0}}'\}$ as the private key held by U_0, where $M_t' = \sum_{a_t} M_1^{c_1} M_2^{c_2} \cdots M_\tau^{c_\tau}$ $(1 \leq t \leq \tau)$ and $M_i' = \sum_{b_i} M_{\tau+1}^{c_{\tau+1}} M_{\tau+2}^{c_{\tau+2}} \cdots M_{n-k}^{c_{n-k}}$ with $a_t = (c_1, \cdots, c_\tau) \in \{0, 1\}^\tau$ and $b_i = (c_{\tau+1}, \cdots, c_{n-k}) \in \{0, 1\}^{n-k-\tau}$.

2.2 Broadcasting of the Messages

This phase corresponds to the actual broadcasting algorithm QB_{K_s} to broadcast the k-qubit message $|P\rangle$ with a suitable state $|B\rangle$ by exploiting a SQC $\mathcal{Q}$.

Step 1. By making use of the generators $M_1, \cdots, M_{n-k}$ and K_s, Alice encodes $|P\rangle$ into $|P\rangle_L$ according to Eq.(1).

Step 2. Alice creates the broadcasted state $|B\rangle = E_B|P\rangle_L$, where E_B is an arbitrary operator with the weight, at most, t. Then, she sends $|B\rangle$ to U_0.

We note that state $|P\rangle$ is encoded by SQC before the generations of the broadcasting state instead of giving $|P\rangle$ directly to tolerate more errors introduced by the possible attackers and the noisy channel. By this way, the messages are broadcasted securely because the state distributing process is also be used for the proof of the unconditional security of QKD by Lo and Chau [6].

2.3 Deciphering of the Broadcasting State

A deciphering algorithm QD_{K_v} is developed here such that Bob is enabled to gain Alice's secret messages.

Step 1. After U_0 have received the broadcasted state $|B'\rangle$, which may be different from $|B\rangle$ because of the effects of the noises and the possible attackers U_0 obtains the operator E'_B from the error syndromes computed by utilizing K_{U_0} and K_v together.

To check the attackers, they may do as follows. when Alice announces E_B publicly, they implement the process of the identity verification. If $E'_B = E_B$, it implies that the qubits of $|B\rangle$ are kept well without being destroyed by the noises. Then, Bob goes to Step 2 to decipher it to obtain the secret messages. Otherwise, some qubits of $|B\rangle$ are destroyed by the noisy channel or the attackers.

Step 2. By performing E'_B onto the received state $|B'\rangle$ from the sender, U_0 obtain $|P\rangle_L = E'_H|B'\rangle$. Then, U_0 have the secret messages $|P\rangle$ by using the decoder of SQC, i.e., performing the inverse process of the operations in Eq.(1).

3 Security Analyses

Firstly, the broadcasted state $|B\rangle$ can not be forged by the attackers or the users in U. It is stipulated that the attacker may be a legitimate user in U_0. It is shown that the transformation from K_{U_0} and K_v to the enciphering session key K_s is intractable. Suppose an attacker can make use of Grover iteration to search for the private key on a quantum computer [7]. Though the algorithm has remarkable efficiency, the attackers must be required, at least, $O(\sqrt{2^{2n}/k})$ times of the search algorithm to catch k operators close to the enciphering session key $\mathcal{X}_1, \cdots, \mathcal{X}_k$ and $O(\sqrt{2^{2n}})$ times of the search algorithm to catch an operators close to E_B. Thus, the broadcasted state $|B\rangle$ of the messages $|P\rangle$ would not be forged by anyone else using the public parameters K_v.

Secondly, it is impossible for the attacker to obtain the distributed messages from the known-plaintext $|B'\rangle$. In this case, the attackers belong to the users out of U_0. To gain the secret messages, they try to obtain the key K_{U_0}, which is kept secret by the legitimate users. However, it is intractable to gain K_{U_0} even if the attacker utilize the Grover iteration to search for them. By making use of the powerful quantum computer, the attackers must require at least, $O(\sqrt{2^{2n}/k})$ times of the search algorithm to catch k operators close to K_{U_0}. According to

the fundamental properties of the quantum physics, the attackers can not obtain the operator E_B by making use of the generators, neither can they decode $|P\rangle$ from $|B'\rangle$. Furthermore, they may also be detected by U_0 in the process of the identity verification.

Finally, it is impossible for Alice to disavow the legally broadcasted messages because the enciphering key K_s have been contained in the broadcasted state $|B\rangle$. If the sender and U_0 are engaged in the disputations because of the sender's disavowal, they need send the broadcasted state $|B\rangle$ and the messages $|P\rangle$ to the arbitrators. If the broadcasted state can be obtained by using the sender's secret key and the messages, the broadcasted state must be carried out by the sender, otherwise, the broadcasted state is forged by the illegal users out of U_0 or the possible attackers.

4 Conclusion

In this paper, we have proposed the locking concept and a secure lock constructed from the generators of the stabilizer of the SQC. By using this secure lock, we have proposed a secure broadcasting scheme for solving the secure broadcasting problem. It is clear that the proposed scheme is a quantum message distribution scheme since each legal user in U_0 can compute the lock by himself. In the end, the proposed scheme is proved to be unconditionally secure.

References

1. Murao M., Macchiavello C., Perinotti P.: Phys. Rev. Lett. **95** (2005) 060503.
2. Murao M., Plenio M. B., Vedral V.: Phys. Rev. A **61** (2000) 032311.
3. Wootters W. K., Zurek W. H.: Nature, **299** (1982) 802-806,.
4. Nielsen M. A.: Quantum computation and quantum information, Cambrige university press (2001).
5. Guo Y., G. Zeng: J. Phys. Soc. Jap., **74**(11) (2005) 2949-2956.
6. Lo H. K., Chau H. F.: Science, **283** (1999) 2050-2056.
7. Grover K.: Phys. Rev. Lett., **79**(2)(1997) 325-329.

Simulation of Quantum Open-Loop Control Systems on a Quantum Computer

Bin Ye[1] and Wen-Bo Xu[1]

School of Information Technology, Southern Yangtze University
No.1800, Lihu Dadao, Wuxi, Jiangsu 214122, P.R. China
yebinxie@yahoo.com.cn,
xwb@sytu.edu.cn

Abstract. A quantum algorithm which allows one to simulate the quantum open-loop control systems is presented. We study the model of steering a spin 1/2 particle via externally electromagnetic fields in detail. Simulation results obtained from the quantum circuit confirm the validity of our approach.

1 Introduction

As Feynman emphasized, quantum computer can efficiently simulate the evolution of certain quantum systems [1,2]. Such systems are computationally very hard for classical simulations as the size of the Hilbert space grows exponentially with the number of particles. More recently, a few quantum algorithms simulating quantum maps were developed to study the complex behavior of the corresponding quantum dynamics. Examples include the baker map [3], the quantum kicked rotator [4] and the quantum tent map [5].

In this paper, we show that quantum simulation can be efficiently used to study quantum control systems. We demonstrate it on the model of manipulating a two-level quantum system which has many important applications in quantum information processing. The advantage of quantum simulation is that one only needs few qubits to simulate the evolution of such open-loop control systems. This allows one to investigate the quantum control problems on the existing nuclear-magnetic-resonance(NMR) quantum computers.

2 System Model

The problem of controlling systems in quantum regime has recently attracted a lot of attention [6,7,8]. In this paper we will take as our model a spin-1/2 particle placed in an external electromagnetic field [8].

The time evolution of the controlled two-level system is determined by the unitary operator $\boldsymbol{U}(t)$, which is formulated in a bilinear equation

$$\dot{\boldsymbol{U}}(t) = \left(\boldsymbol{H}_0 + \sum_{k=x,y,z} \boldsymbol{H}_k u_k(t) \right) \boldsymbol{U}(t) \tag{1}$$

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 45–48, 2006.

where the internal Hamiltonian $\boldsymbol{H}_0$ and the external Hamiltonian $\boldsymbol{H}_k$ are Hermitian linear operators, and $u_k(t)$ are the components of the external electromagnetic fields which play the role of controls. The optimal control problem for (1) is defined as: for a specified final time T and a target final matrix $\boldsymbol{U}_f$, find a set of control inputs u_x, u_y, u_z such that $\boldsymbol{U}(T, u_x, u_y, u_z) = \boldsymbol{U}_f$ and the energy function

$$J(u) = \int_0^T \sum_{k,j=x,y,z} q_{kj} u_k(t) u_j(t) \mathrm{d}t \tag{2}$$

is minimized with the form $\{q_{kj}\}$ being positive-definite.

Using algebraic dynamical method, the exact analytical solution for the optimal control problem can be obtained. In the case of two component of the electromagnetic field being varied, say u_x and u_y, the optimal solutions for the problem of flipping the spin have the form $u_x^e(t) = a\cos(\omega t + \phi)$ and $u_y^e(t) = a\sin(\omega t + \phi)$, where a, ω and ϕ are constant parameters. As a result, the possible minimum cost of the energy function (2) is:

$$J(u) = (2n+1)^2 \frac{\pi^2}{4T} \qquad n = 0, 1, 2, \cdots \tag{3}$$

and the corresponding evolution operator is

$$\boldsymbol{U}(t) = \begin{pmatrix} e^{-i(\frac{\omega t+\phi}{2})} & 0 \\ 0 & e^{i(\frac{\omega t+\phi}{2})} \end{pmatrix} \begin{pmatrix} \cos at & -i\sin at \\ -i\sin at & \cos at \end{pmatrix} \begin{pmatrix} e^{i\phi} & 0 \\ 0 & e^{-i\phi} \end{pmatrix} . \tag{4}$$

3 Quantum Simulation Analysis

An evolution operator $\boldsymbol{U}$ can generally be described by some self-adjoint operator $\boldsymbol{H}$ to yield $\boldsymbol{U} = e^{-i\boldsymbol{H}}$ (we set $\hbar = 1$ throughout the paper). To simulate the unitary operator $\boldsymbol{U}$ with a finite dimension $N = 2^{n_q}$, we will represent the state $|\psi\rangle$ with a quantum register of n_q qubits. Then we transform the eigenstates $|n\rangle$ of $\boldsymbol{H}$ to binary notation and identify them with the quantum register states

$$|\alpha_0, \alpha_1, \cdots, \alpha_{n_q-1}\rangle = |\alpha_0\rangle_0 |\alpha_1\rangle_1 \cdots |\alpha_{n_q-1}\rangle_{n_q-1} \tag{5}$$

where $\alpha_j \in \{0, 1\}$ and the states $|0\rangle_j$ and $|1\rangle_j$ are the two basis states of the jth qubit. Thus for a diagonal evolution operator $\boldsymbol{U}(t)$, it can be expressed in terms of elementary quantum gates by:

$$\begin{aligned} \boldsymbol{U}(t)|\psi\rangle &= e^{-it\boldsymbol{H}}|\psi\rangle \\ &= e^{-itn}|n\rangle \\ &= \prod_{j=0}^{n_q-1} \boldsymbol{B}_j^{(1)}(-t \cdot 2^j)|\alpha_0 \alpha_1 \ldots \alpha_{n_q-1}\rangle \end{aligned} \tag{6}$$

where $\boldsymbol{B}_j^{(1)}(\phi)$ is the phase-shift quantum gate which provides a phase factor $e^{i\phi}$ if $\alpha_j = 1$.

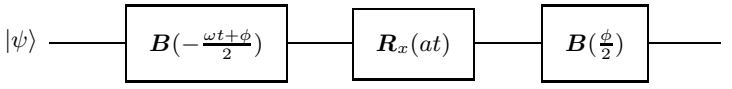


Fig. 1. Quantum circuit simulating the optimal control of a spin-1/2 particle. The one-qubit quantum gate $e^{-i(at)\sigma_x}$ is denoted by $\boldsymbol{R}_x(at)$ which is actually an X-axis rotation operation

Recalling the evolution operator (4) in the optimal control model, the first and third factors on the right side of the equation can be realized solely in terms of the phase-shift gates according to the mechanism described above. As regards the second factor, we will use a one-qubit quantum gate $e^{-i(at)\sigma_x}$ to implement it in which σ_x is the Pauli spin matrix. This approach is based on the following theorem:

Theorem 1. *Suppose* $\boldsymbol{A}$ *is a matrix satisfying* $\boldsymbol{A}^2 = \boldsymbol{I}$, $x \in \mathbb{R}$, *then the following equation holds:*

$$e^{i\boldsymbol{A}x} = \cos(x)\boldsymbol{I} + i\sin(x)\boldsymbol{A} \ . \tag{7}$$

Proof. Expand $\cos \boldsymbol{A}$ and $\sin \boldsymbol{A}$ as the series $\cos \boldsymbol{A} = \sum_{n=0}^{\infty} \frac{1}{(2n)!}(-1)^n \boldsymbol{A}^{2n}$ and $\sin \boldsymbol{A} = \sum_{n=0}^{\infty} \frac{1}{(2n+1)!}(-1)^n \boldsymbol{A}^{2n+1}$. Substituting $e^{i\boldsymbol{A}x}$ by the series and following $\boldsymbol{A}^2 = \boldsymbol{I}$, one can easily verify the equation (7). □

A quantum circuit simulating the time evolution of the quantum state in the optimal control system is provided in Fig. 1. Note that the sequence of simple matrix factors to accomplish the evolution operator $\boldsymbol{U}(t)$ is by no means unique [7]. Using the procedure above, however, similar quantum algorithms can be developed for other constructive decompositions.

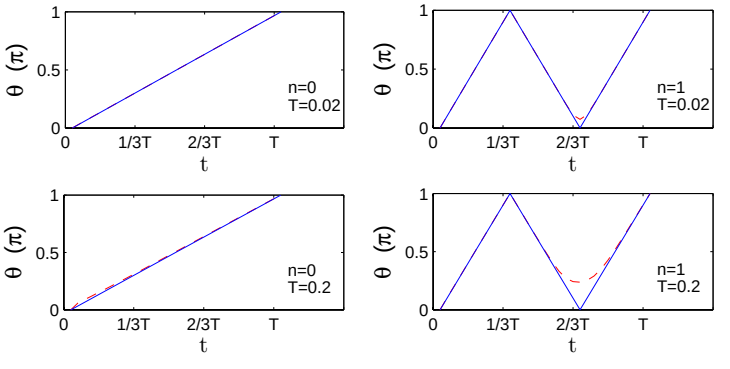


Fig. 2. Time-variation of θ in the process of flipping a spin-1/2 particle. The full line in each figure corresponds to the quantum simulation while the dashed lines are for the case of classical simulation. Various time period T and number of n in (3) are represented in the lower-right corner of each figure.

4 Simulation Results

In order to show the quantum state evolution explicitly, it is mapped onto a Bloch sphere. A point on Bloch sphere describes a qubit state $|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle$, with $\theta \in [0, \pi], \varphi \in [0, 2\pi]$. Since the factor $e^{i\varphi}$ is of no effect on the probability amplitudes, we will pay our attention only to the variation of θ.

Several simulations are performed with various control inputs. Simulation results obtained by the proposed quantum algorithm are compared with that of classical simulation. The frequencies in the control functions are choose as the resonance frequency $\omega = 2$. For simplicity, the initial phases in the control functions are set to $\phi = 0$. The results are shown in Fig. 2. As expected, the quantum simulation tracks the trajectory of classical simulation with little or no error.

5 Conclusions

There has been significant interest in controlling the two-level quantum systems. We have developed a quantum algorithm to efficiently simulate the optimal control model using few qubits. The quantum simulation results are in agreement with that of classical simulation with little or no error. The algorithm presented here shows that quantum open-loop control systems can be simulated efficiently on a quantum computer.

References

1. Feynman, R.P.: Quantum Mechanical Computers. Foundations of Physics **16** (1986) 507–531
2. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press (2000)
3. Schack, R.: Using a Quantum Computer to Investigate Quantum Chaos. Physical Review A **57** (1998) 1634–1635
4. Levi, B., Georgeot, B., Shepelyansky, D.L.: Quantum Computing of Quantum Chaos in the Kicked Rotator Model. Physical Review E **67** (2003) 46220–46229
5. Frahm, K.M., Fleckinger, R., Shepelyansky, D.L.: Quantum Chaos and Random Matrix Theory for Fidelity Decay in Quantum Computations with Static Imperfections. The European Physical Journal D **29** (2004) 139–155
6. Doherty, A.C., Habib, S., Jacobs, K., Mabuchi, H., Tan, S.M.: Quantum Feedback Control and Classical Control Theory. Physical Review A **62** (2000) 12105–12117
7. Ramakrishna, V., Flores, K.L.: Quantum Control by Decompositions of SU(2). Physical Review A **62** (2000) 53409–53419
8. D'Alessandro, D., Dahleh, M.: Optimal Control of Two-Level Quantum System. IEEE Transactions on Automatic Control **46** (2001) 866–876

An Optimization Algorithm Inspired by Membrane Computing*

Liang Huang and Ning Wang**

National Laboratory of Industrial Control Technology, Institute of Advanced Process Control, Zhejiang University, Hangzhou, 310027, P.R. China
{lhuang, nwang}@iipc.zju.edu.cn

Abstract. Inspired by Membrane Computing, this paper presents an optimization algorithm for the problems with large feasible space and large parameters. The algorithm consists of two parts. One is to determine the good area and another is a fast searching tool.The comparisons among some evolutionary algorithms illuminate that the algorithm is simple and computes fast with high precision on optimal problems.

Keywords: Optimization, membrane computing (P system), function optimization, genetic algorithm, evolutionary algorithm.

1 Introduction

Many complex problems are very difficult to optimize. For the intractable problems, this paper presents a new algorithm. Inspired by Membrane Computing [1,2], the algorithm consists of two parts. One adjusts the good area, where the optimal solution exists with high probability. Another part is a cell-like fast searching tool which searches the best solution. The algorithm repeats the modification of the best area and optimal solution until the stopping condition is satisfied.

2 The Construction of the New Algorithm

The fast searching tool is the most important part of the algorithm. It adopts the similar structure (figure 1) of Membrane Computing [1,2]. In each region of the tool, there are some chromosomes. They evolve as the rules, which belong to the regions exist respectively. It has the evolution rule, the crossover rule and the communication rule. The crossover rule is same as that in genetic algorithm. In a region, one chromosome S evolves into S' according to the following evolution rule.

$$\begin{cases} S \to S' \\ S = x_1 x_2 \ldots x_l \\ S' = y_1 y_2 \ldots y_l \end{cases} \tag{1}$$

* This work is supported by the National Natural Science Foundation of China 60421002.

** To whom correspondence should be addressed.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 49–52, 2006.

where $y_i = x_i$ or $y' = x_i + \eta_i$, $i = 1, 2, \ldots, l$, x_i, y_i are the variables of a problem.The chromosome S is a feasible solution of the problem, which are coded by l variables. η_i is the range of distribution which is different in different membranes.

The communication rules means that the membrane sends some maximum strings into the outer membrane which directly contains it. The rule can be written as follows.

$$[a_{max1}, a_{max2}, \ldots, a_{maxn}]_i \rightarrow []_i a_{max1}, a_{max2}, \ldots, a_{maxn} \tag{2}$$

$[]_i$ represents the membrane i. $a_{max1}, a_{max2}, \ldots, a_{maxn}$ are n maximum strings in the region i . The tool converges very fast because of the communication between membranes.

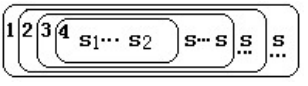


Fig. 1. The structure of the searching tool

It is difficult to use some evolutionary algorithms to deal with the problems with a large search space. The algorithm can select a small good area for the fast searching tool. First, it randomly collects samples in different areas of the whole feasible space. Second, the area with a low mean fitness of samples will be deleted, or the area with a high mean fitness of samples will remain as the good area. The system objective function is expressed as follows.

$$\text{Minmize:} \quad y = f(x_1, x_i, \ldots, x_l). \tag{3}$$

where $x_i \in [u_{ilow}, u_{iup}]$, u_{iup} is the upper limit of range of x_i, u_{ilow} is the lower limit of range of x_i . u_{im} is the mean value of range of x_i ; For example, the feasible space of each variable is divided into two parts. Then, the whole feasible solution of the problem will be divided into 2^l subspaces. We adopt random n chromosomes to represent dots in the feasible space. Let y_j denote a function value of the chromosome j, where $j = 1, \ldots, n$. Suppose there are n_1 chromosomes, which gene $x_i \in [u_{im}, u_{iup}]$ and there are n_2 chromosomes, which gene $x_i \in [u_{ilow}, u_{im}]$. Definitions of the mean value of areas mf_{i1}, mf_{i2} are as follows.

$$\begin{cases} mf_{i1} = \frac{1}{n_1} \sum_{h=1}^{n_1} y_h \\ mf_{i2} = \frac{1}{n_2} \sum_{k=1}^{n_2} y_k \end{cases} \tag{4}$$

If $mf_{i1} < mf_{i2}$,the area $[u_{ilow}, u_{im}]$ will be deleted and the area $[u_{im}, u_{iup}]$ will be selected as the good area. On the contrary, if $mf_{i1} > mf_{i2}$, the area $[u_{ilow}, u_{im}]$ will be selected as the good area.

If the best solution is found out of the good area but in the global feasible space, the area will be enlarged to comprise the adjacent field of the best solution. Then, the algorithm randomly produces new initiate chromosomes for the searching tool from the new good area. However, the best solution of the last generation is kept in the new population of the fast searching tool.

3 Performance Comparisons

The performance of the new optimization algorithm is compared with some evolutionary algorithms using benchmark functions as following table 1 [3]. The average results of 30 independent runs are listed in Table 2. The experimental results of the algorithms come from [3]. In table 2, IEA refers to the intelligent evolutionary algorithms [3]. OGA is an orthogonal genetic algorithm [4]. Other algorithms with elitist strategy and the associated crossover are denoted as OEGA (one-point), TEGA (two-point), UEGA (uniform) and BLXGA (BLX-a). In the experiments, each parameter is encoded using 10 bits for all test functions except that of function f_1 which uses 24 bits. For all compared evolutionary algorithms, the populations are 10 and the stopping condition is 1000 generations. There are 25 chromosomes in each region of the fast searching tool of the new algorithm. In other words, the population is 100. In addition, 5% of the good area will be deleted when the fast tool runs 10 generations. The results are shown in following table 2 when the algorithm evolves 50 generations and 5000 generations. To illustrate the performance of comparisons on various numbers of parameters as following figure, the distance $dist(D)$ is used to describe the mean distance between the optimal solution $f_{opt(D)}$ and the obtained best solution $f_{best(D)}$for one parameter as follows.

$$dist(D) = \frac{|f_{opt}(D) - f_{best}(D)|}{D} \tag{5}$$

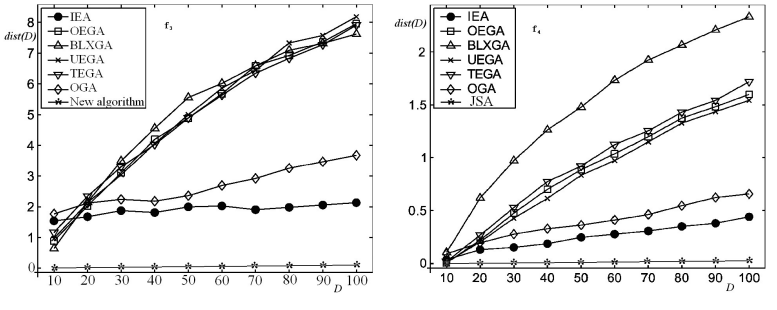


Fig. 2. $dist(D)$ curves for f_3 and f_4

Table 1. Benchmark Functions [3]

Test function	Domain	Optimum(minize)
$f_1 = \frac{1}{4000}\sum_{i=1}^{D} x_i^2 - \prod_{i=1}^{D}\cos(\frac{x_i}{\sqrt{i}}) + 1;$	$[-600, 600]$	0
$f_2 = \sum_{i=1}^{D-1}[100(x_{i+1} - x_i^2 + (x_i - 1)^2];$	$[-5.12, 5.12]$	0
$f_3 = \sum_{i=1}^{D}[x_i^2 - 10\cos(2\pi x_i) + 10];$	$[-5.12, 5.12]$	0
$f_4 = 6D + \sum_{i=1}^{D} floor(x_i);$	$[-5.12, 5.12]$	0

Table 2. Mean fitness of functions with 10 dimensions and 100 demensions

function(D)	IEA	OEGA	UEGA	TEGA	BLXGA	OGA	MC(50G)	MC(5000G)
$f_1(10)$	0.999	1.001	1.030	1.002	1.030	1.002	0.334	0.0152
$f_2(10)$	116.4	41.4	40	23	22.6	94.6	14.6	0.27
$f_3(10)$	15.4	8.9	9.5	11.6	6.5	5.3	4.2	1.2×10^{-10}
$f_4(10)$	0.338	0.03	0.067	0.17	1.03	0.9	3.8	0
$f_1(100)$	32.86	511.93	1002.00	537.09	237.84	48.25	29.57	0.0006
$f_2(100)$	2081	96556	89514	97990	16086	5282	1107	93
$f_3(100)$	211.3	795.7	819.6	791.5	763.4	367.5	474.4	0.43
$f_4(100)$	43.9	159.9	154.4	172	233.6	65.2	269.7	0

4 Conclusions

This paper proposed a new evolutionary algorithm for solving optimization problems with large feasible solution space and large parameters. It is shown empirically that the optimization algorithm has good performance in solving benchmark functions as compared with some existing evolutionary algorithms. Due to its simplicity, converge fast,theoretical elegance, generality, and superiority, the optimization algorithm can be used to solve complex problems.

References

1. Gh. Păun. Computing with membranes. Journal of Computer and System Sciences, 61,1(2000),108-143.
2. Gh. Păun, From Cells to Computers: Computing with Membranes (P Systems), Proc. Intern. Workshop Grammar Systems 2000 (R. Freund, A. Kelemenova, eds.), Bad Ischl, Austria, July 2000, 9-40, and BioSystems, 59, 3 (2001), 139-158.
3. Shinn-Ying Ho, Jian-Huang Chen. Intelligent Evolutionary Algorithms for Large Parameter Optimization Problems. IEEE Transactions on Evolutionary Computation, 8,6(2004),522-541.
4. Y.W.Leung and Y.P. Wang, An orthogonal genetic algorithm with quantization for global numerical optimization. IEEE Transactions on Evolutionary Computation, 5,1(2001),41-53.

A Mapping Function to Use Cellular Automata for Solving MAS Problems

Andreas Goebels

International Graduate School of Dynamic Intelligent Systems
Knowledge Based Systems, University of Paderborn, Germany
swarmgroup@upb.de

Abstract. Cellular automata are a very powerful and well researched area in computer science. We use approaches from the cellular automata research to solve optimization problems in the multi agent system research area. For this purpose, we require a transformation from agents located in an Euclidean space into an abstract cell assignment for cellular automata. In this paper, a mapping function is presented and evaluated with a reverse function. This function can be calculated by each agent individually based only on local information. Additionally, we examine the performance of the function in inexact and non-deterministic environments.

1 Introduction

Cellular automata (CA) are a very powerful and well researched area in computer science and are able to cope with a lot of problems. In [1] we used a genetic algorithm to create suitable sets of rules for cellular automata that could successfully solve an optimization problem (the *Online Partitioning Problem*) from the area of multi agent systems (MAS). The ideas for this approach came from classification problems in CAs. The *Majority Classification (MC)* or *Density Classification (DC)* task is a well known and exhaustively studied problem in the field of Cellular Automata [2], first published by Gacs, Kurdyumov and Levin [3]. There, an one-dimensional binary CA with an initially random state for each cell should classify the initial assignment depending on the ratio of 0's to 1's. This means, if there were more 1's than 0's in the input, the final state of each cell should be $'1'$, if there were more 0's, the final state of each cell should be $'0'$. We define the final state to be the CA assignment after a predefined and fixed number of transitions.

In our approach we used several ideas from the field of the majority classification task for two-dimensional CAs to solve the *Online Partitioning Problem* (OPP) introduced in [4]. One key restriction of the *OPP* is the absence of any kind of central instance, all decisions of the agents are based only on local information, received by communication with direct neighbours. Thus, CAs seem to be a good concept for this kind of problems because of the exclusive interaction with cells that are located nearby. Therefore, we incorporate the ideas of Mitchell et al. [5] and transfer them from one dimensional CAs to two dimensional ones. The activated cells in the cellular automata represent the agents and the value of an activated cell describes the currently chosen target according to the *OPP* .

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 53–62, 2006.

One open issue in [1] was the mapping from agents, located in an Euclidean space, to a CA assignment. This mapping should take the neighbourhood properties of the agent set into account. There exist several ideas to embed graphs in data structures or grids (e.g. [6]), but these operate only with the whole graph as input. Here, we present a mapping function in a deterministic and a random version, based only on local information. Additionally, we limit the agents' abilities by introducing inexactness and limited view radii. Hence, our approach is suitable for real world applications with uncertain sensor inputs.

The paper is organized as follows. Section 2 gives an overview of the researches in the Majority/Density Classification Task for one-dimensional cellular automata and the Online Partitioning Problem, which is formally defined there. In section 3 the mapping function is presented and motivated and its quality is shown in experimental runs. In the next section 4 we introduce inexactness and limited agents' visibility and in the end we conclude the paper and give ideas for future research.

2 Related Work

The *Majority Classification (MC)* or *Density Classification (DC)* task is a well known and exhaustively studied problem in the field of Cellular Automata (CA) [2], first published by Gacs, Kurdyumov and Levin [3]. A one-dimensional binary CA with an initially random state for each cell $c_i \in \{0, 1\}$ should result in a uniform state for all cells depending on the initial ratio of 0's to 1's. This means, if there were more 1's than 0's in the initial configuration, the final state of each cell should be $'1'$, if there were more 0's, the final state of each cell should be $'0'$. We define the final state to be the CA assignment after a predefined and fixed number of transitions.

There exist many different approaches to generate rules for classifying a high proportion of such random initial cell assignments. The best approaches classify more than 86% correct. Manually generated rules reach a correct classification value of around 82%; the best one has been obtained by Das, Mitchell and Crutchfield with 82.178%.

For the automatic generation of rules for cellular automata, several techniques have been used, e.g. genetic algorithms [5], genetic programming [7] or coevolution [8][9] [10]. In [11] an exhaustive study of the *Majority Classification Task* has been made. There, Kirchner could reproduce most of the results obtained by Mitchell et al. Under our supervision she developed a software tool that can also be used to group the values in a randomly initialized one-dimensional cellular automata.

The Online Partitioning Problem (*OPP*) is a multi objective optimization problem dealing with the distribution of agents with very limited and mostly local knowledge onto different tasks represented as targets. The agents distribute themselves autonomously according to the following three objectives:

(1) The agents have to be distributed uniformly
(2) Minimize the overall distance toward the targets
(3) The abilities of the agents should be very simple

Each of these goals is oppositional to any other, so we look for the best possible solution fitting in all objectives in quite an acceptable way. We can put this objectives

in a more formal way by defining the first two objectives mathematically: An instance of *OPP* with $OPP = (\mathcal{A}, \mathcal{T}, \rho)$ with the agent set $\mathcal{A} = \{a_1, ..., a_n\}$, the target set $\mathcal{T} = \{t_1, ..., t_m\}$ and a location function ρ is an optimisation problem. $\delta(p_1, p_2)$ defines the geometric distance between two points in the Euclidean space. This function works in the same way if we consider two arbitrary agents a and a' or an agent a_k and a target t_l, then $\delta(a, a')$ and $\delta(a_k, t_l)$ calculate the distance between the positions of two agents or between an agent and a target, respectively.

$$\delta : \mathcal{A} \text{ x } \mathcal{A} \rightarrow \mathbb{R} : \delta(a, a') = d \text{ and } \delta : \mathcal{T} \text{ x } \mathcal{T} \rightarrow \mathbb{R} : \delta(t, t') = d$$

$$\delta : \mathcal{A} \text{ x } \mathcal{T} \rightarrow \mathbb{R} : \delta(a, t) = d = \delta : \mathcal{T} \text{ x } \mathcal{A} \rightarrow \mathbb{R} : \delta(t, a) = d$$

The question is whether there are subsets $\mathcal{S} = \{S_1, S_2, ..., S_k\}$ with $\mathcal{A} = \{S_1 \cup S_2 \cup ... \cup S_k\}$ such that the following objectives are minimal:

$$-\prod_{i=1}^{|\mathcal{S}|} |S_i| \text{ and } \sum_{i=1}^{|\mathcal{S}|} \sum_{a \in S_i} \delta(a, t_i)$$

These can be combined by the following formula:

$$f = \alpha \cdot \left(\frac{\prod_{i=1}^{m} b_i}{\prod_{i=1}^{m} o_i} \right) + \beta \cdot \left(\frac{\sum_{i=1}^{n} \min_{j=1..m} (\delta(a_i, t_j))}{\sum_{i=1}^{n} \delta(a_i, target(a_i))} \right)$$

with

$$\alpha + \beta = 1; \alpha, \beta \geq 0$$

f has to be maximized. $b_i(o_i)$ is the current (optimal) number of agents that have chosen target i and $target(a)$ specifies the target the agent a has chosen. All agents are able to communicate with their direct neighbours and know the order of the targets according to their distance. A more detailed description of the abilities of the agents can be found e.g. in [4].

The *Online Partitioning Problem* belongs to the area of Multi Agent Systems and Swarm Intelligence. Swarm Intelligence deals with large sets of individuals or agents that can be seen as a self organizing system showing emergent behaviour [12][13]. Several basic strategies have been presented and we compared our results obtained with the CA-approach in [1] with those obtained in [4]. The difference between this new CA based approach and the earlier published ideas is, that here the CA rules only guide the decision process of the agents for a target in the first few time steps.

3 Mapping Function

Our idea presented in this paper is to map a set of agents, that are located somewhere in space, onto a two-dimensional cell-grid with two main states, *active* and *inactive*; an active grid field represents the presence of an agent. This grid is the cell assignment of a cellular automata. The transformation from an agents' positions to the grid will

be realized by a mapping function, denoted by ψ. It is important to carry-forward the neighbourhood relations of agents in the space to the activated cells in the cellular automata. Hence, agents that are located nearby in space should be direct neighbours in the grid. One major problem for such a function ψ is the decision if an agent is closeby, because this decision has to be calculated based on exclusively local information. ψ has to decide for each agent $a \in \mathcal{A}$ with a fixed but only locally known position p in space, how the adjacent $\mathcal{N} = \{n_0, n_1, ..., n_8\}$ cells (including the center cell n_0 itself representing the current agent) on the CA should look like to represent the original situation as best as possible.

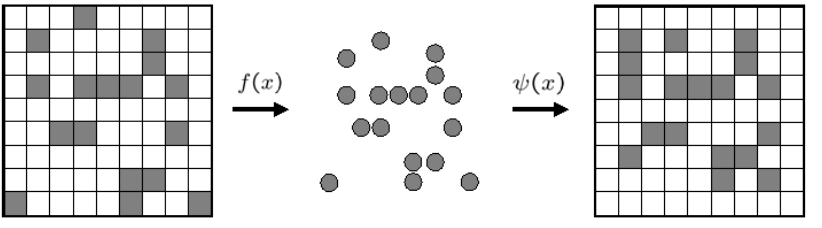


Fig. 1. How to measure the mapping-quality

To test the quality of such a function ψ, we need the ideal mapping as a reference value. Therefore, we decided to start with a random cellular automata $\mathcal{CA}$ and a mapping function f from activated cells of a CA to coordinates of agents in space. This agent set is the input for our mapping function ψ. The CA description $\mathcal{CA}_{mapped}$, that can be derived from ψ, can then be compared with $\mathcal{CA}$.

The function $f : \mathcal{CA} \rightarrow \mathbb{R}\text{x}\mathbb{R}$ can be defined by

$$f(c_{i,j}) = \begin{cases} \text{agent on position } (i,j) \text{ if } c_{i,j} \text{ is activated in } \mathcal{CA} \\ \text{no agent on position } (i,j) \qquad\qquad else \end{cases}$$

The idea of our mapping function ψ is to find for each agent a an appropriate representation for the terms 'near' and 'far' and to decide for each nearest agent in one of the eight sectors around a if it is 'near' or 'far'. The first problem is solved by the consideration of the nearest and the farthest among the eight closest agents in the sectors. These two extreme values constitute our normative system. The second part is realized by using a special gradient for ψ. The concept to divide the space into sectors originating at each agent is similar to the construction of efficient graphs for mobile ad hoc networks ($MANETS$). One example of such successful graphs is a special kind of Yao graphs, the sparse Yao graph (SparsY, see [14] and [15]). For these graphs, each node divides the space into k cones and draws an edge to the nearest node in each sector. Therefore, the in- and the out-degree of each node is limited to at most k.

3.1 Deterministic Mapping

We start with the mapping of the (highly environment distance measurement dependent) distances from the agent a to its neighbours in each sector into the standardised

interval $[0;1]$. After the mapping, the distance value of the closest agent is 0 and the distance value of the agent that is farthest away is 1. This distance values are now totally independent from the original values, so we do not longer have to take into consideration if we are dealing with kilometres or nanometres etc. The conversion is done by the formula

$$\text{r-dist}(a,a_k) = \frac{\delta(a,a_k) - \min\limits_{i=1,\ldots 8}\{\delta(a,a_i)\}}{\max\limits_{i=1,\ldots 8}\{\delta(a,a_i)\} - \min\limits_{i=1,\ldots 8}\{\delta(a,a_i)\}} \in [0;1] \tag{1}$$

for each agent separately depending on the nearest agent in each sector[1]. In this formula, the agent a_i is the nearest agent in the sector i, this sector can be described as the cone with an angle of $\left(\frac{2\cdot\pi}{8}\right)$ with sides from $\left(\frac{i-1}{8}\cdot 2\pi\right)$ to $\left(\frac{i}{8}\right)\cdot 2\pi$.

Next, we define a strictly monotonic decreasing function ψ.

$$\psi(x) = 2^{-\frac{x^2}{\left(\frac{1}{8}\cdot\sum\limits_{i=1}^{8}\text{r-dist}_{(a,a_i)}\right)^2}}$$

This is our mapping function that determines for an arbitrary relative distance (r-dist) value $x \in [0;1]$ if this is a far or a near distance. The decision, if an agent is neighbour or not, depends on the value of this function, and we can decide for each sector if the agent is close enough to appear in the neighbourhood of agent a or not. Therefore, the value for each cell $n_0,\ldots,n_8$ in the neighbourhood of an arbitrary agent a is defined by

$$n_k = \begin{cases} active & \text{for } k=0; \\ active & \text{for } \psi(\text{r-dist}(a,a_k)) \geq \frac{1}{2}; \\ inactive & else \end{cases}$$

3.2 Random Mapping

The random mapping is a slight modification of the deterministic mapping presented before. We allocate the adjacent cells of agent a with a probability that depends on the function ψ.

$$n_k = \begin{cases} active & \text{for } k=0; \\ active & \text{with probability } p=\psi(\text{r-dist}(a,a_k)); \\ inactive & else \end{cases}$$

3.3 Measuring the Error Rate

We tested our mapping function $\psi(x)$ by comparing the mapping result onto a cellular automata with the optimal solution. Therefore, we start with a cellular automata cell assignment $\mathcal{C}_1$ with a varying density d of activated cells ($d \in \{0\%;\ldots,100\%\}$). The

[1] If the denominator of the fraction is zero, we force r-dist to return 0.5 because all agents have exactly the same distance.

activated CA-cells will be converted into a set of agents with position in space based on the cell position in the CA. For cell c_{ij} the position in space will be (i, j). Then we applied our algorithm to this set of agents and convert them back into a cellular automata cell assignment $\mathcal{C}_2$. The quality of the final assignment $\mathcal{C}_2$ is measured by two error

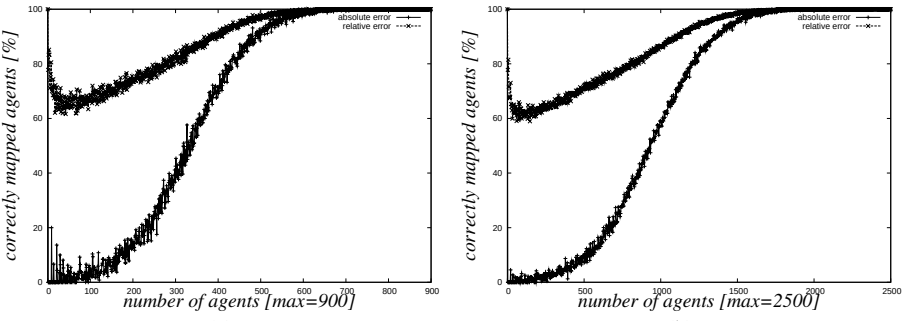


Fig. 2. The development of the absolute and the relative error metrices for different numbers of agents. The curves are nearly identical. This is due to the only local decisions the agents make.

metrics. The absolute error metrix represents the ratio of correspondent activated cells that have an identical neighbourhood in $\mathcal{C}_1$ and $\mathcal{C}_2$ to the overall number of activated cells. The relative error regards similarities in the neighbourhood of correspondent activated cells. In figure 2 we compare the mapping quality of our algorithm for different CA-sizes. The absolute error err_a is calculated by

$$err_a = \frac{\sum\limits_{c \in \mathcal{C}_2 \ with \ c \neq \ empty} \begin{cases} 1 \text{ for } \ n(c \in \mathcal{C}_1) = n(c \in \mathcal{C}_2) \\ 0 \text{ else} \end{cases}}{|\{c | c \in \mathcal{C}_2 \ and \ c \neq \ empty\}|}$$

and the relative error err_r by

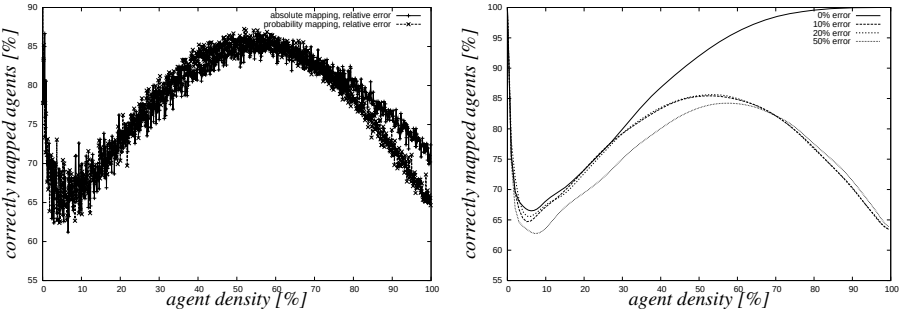
$$err_r = \frac{\sum\limits_{c \in \mathcal{C}_2 \ with \ c \neq \ empty} |\Delta(n(c \in \mathcal{C}_1), n(c \in \mathcal{C}_2))|}{9 \cdot |\{c | c \in \mathcal{C}_2 \ and \ c \neq \ empty\}|}$$

In this formulas we use $n(c)$ as a function representing the Moore neighbourhood of cell c (and c itself) and the difference function $\Delta(n, n')$ calculates the number of differences between the neighbourhoods of n and n' with $\Delta(n, n) \in \{0, ..., 9\}$.

4 Introducing Inexactness and Limitations

In the previous section we assumed that the distances could be measured with no kind of noise and 100% preciceness. Now we consider the behaviour of our algorithm if we add noise to the distance measurements. Therefore, we modify the correct distance value by p percent ($dist_m = dist + m$ with $m \in [-p \cdot dist; p \cdot dist]$). The other parameters

remain the same. Figure 3(a) visualizes that the characteristics of the random and the absolute mapping show a high similarity, but for cell densities greater 50% the quality of the random mapping is significantly better. At the end there is a difference of more then 5%.



(a) (b)

Fig. 3. Comparison of the absolute and the probability mapping function when the agents cannot measure distances exact. In figure 3(a), we used an error of at most $p = 10\%$. In (b), we compare the development for different error sizes.

Following the success of our mapping function in an ideal environment and satisfying results when adding inexactness, we analyse our approach with more realistic agents, i.e. agents with limited sensor information in respect of view radius. In detail, we observed the behaviour of our ψ while applying different view radii. Therefore, we modify our original formula (1) as follows:

$$\text{r-dist}_{vr(r)}(a, a_k) = \frac{\delta_{vr(r)}(a, a_k) - \min\limits_{i=1,\ldots 8}\{\delta_{vr(r)}(a, a_i)\}}{\max\limits_{i=1,\ldots 8}\{\delta_{vr(r)}(a, a_i)\} - \min\limits_{i=1,\ldots 8}\{\delta_{vr(r)}(a, a_i)\}} \in [0; 1]$$

The only difference between the r-dist$_{vr(r)}$ and the prior r-dist function is that we now have two case differentiations:

$$\delta_{vr(r)}(a, a_k) = \begin{cases} \delta(a, a_k) & for \;\; r \gtrsim \delta(a, a_k); \\ \infty & else \end{cases}$$

In the figures 4(a)-4(c) we compare the quality of our mapping function for agents using limited view radii r. The influence of the real value r depends obviously onto the real value distances of the agents in the space. For the experiments presented here we chose the r's from a reasonable interval[2]. In the experiments you can observe that the view radius cannot be chosen in a really wrong way. We distinguish three different cases:

[2] A *reasonable* interval $[I_1; I_2]$ is obviously an interval with the minimum and maximum distance of agents in the simulation space as interval borders. Hence, I_1 should be equal to or smaller as the minimum distance between any two agents a and a' and I_2 should be greater as or equal to the maximum distance between any two agents a and a'.

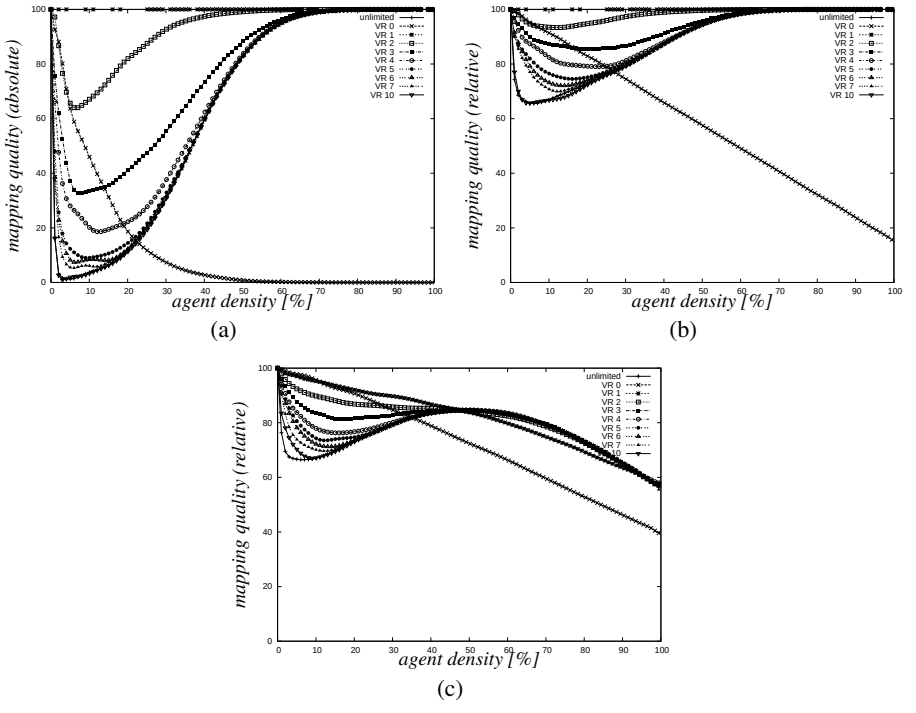


Fig. 4. Comparison of different view radii of the agents. Again, we compare the results of our mapping function for different agent densities. The figures 4(a) and 4(b) show the behaviour of our mapping function without errors in the distance measurement. The absolute and the relative error is shown. Figure 4(c) is the result of increasing the distance inexactness of the agents to 10%.

No Sensing: If the view radius is smaller than the minimum distance of any two agents, the mapping function can obviously map the agent set only into a cellular automata in which no cell has any neighbour. Therefore, the quality of the mapping function is decreasing when the density of the agents and therewith the average number of neighbours in the original CA is increasing. This can be seen in all settings in figure 4 considering the quality development for the view radius $VR = 0$.

Over Sensing: If the view radius is greater than the maximum distance between any two agents, there is no change in the knowledge of the agents because with an increase of the radius no additional information can be gained.

Variable Sensing: If the view radius is inside the reasonable interval then the quality of the mapping function changes. We can distinguish between two cases. For sparse agent densities, a decrease of r increases the mapping quality. In such settings, agents that are far away can be misleadingly considered as near neighbours because there are not enough agents to calculate a good normative distance value base. An example for such a bad case is shown in figure 5. For an agent density of approximately 50% there is no longer a significant difference for the different radii VR. There is one special case if the view radius is equal to the minimal distance

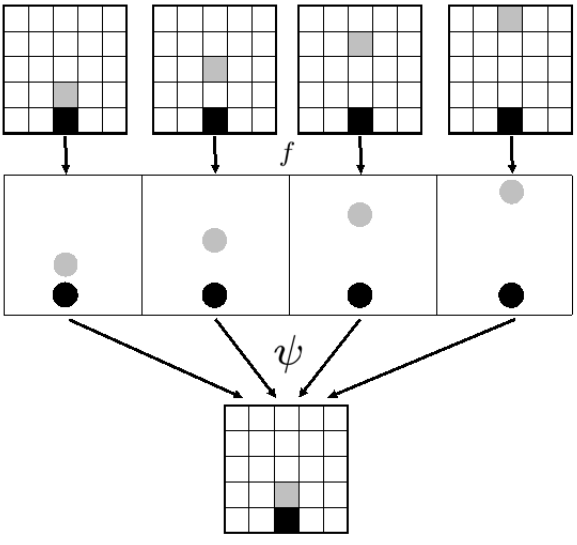


Fig. 5. Due to the missing global information, the mapping function can not distinguish between the four originally different cell assignments. For our formula, the grey agent is the nearest one in all settings and therefore a neighbour for the black agent.

between any two agents. This radius is identical to the function we use to map the agents from the original CA to the agent set in space. In the experiments with no inexactness, $r = 1$ is a direct inversion of this function and therefore the mapping quality is always exact 100%. In the experiments with inexactness, there are some agents that could not be sensed because they are out of the sight radius. Therefore, we receive slightly worse results.

5 Conclusion and Future Work

In this paper, we presented a mapping function to transform a group of agents into active and inactive cells of a cellular automata. The quality of this function has been determined by a comparison with the ideal cell assignment obtained via a double transformation process. To make the system more realistic, we added inexactness and limited agent abilities. In all combinations, our mapping function ψ shows an behaviour from acceptable to very good, mostly depending on the agent density in the system. The best results could be obtained if the agent distribution is not too sparse. Consequently, if the density in the agent setting is below 50%, our approach could only produce a mapping with up to 40% differences to the best mapping. Such worst case scenarios have been presented.

In our future research, we will try do prove the quality of this best approach and present more worst case scenarios.

References

1. Goebels, A., Weimer, A., Priesterjahn, S.: Using cellular automata with evolutionary learned rules to solve the online partitioning problem. In: Proceedings of the IEEE Congress on Evolutionary Computation (CEC'05), Edinburgh (2005) 837–843
2. Panait, L., Luke, S.: Cooperative multi-agent learning: The state of the art. In: Technical Report GMU-CS-TR-2003-1, George Mason University, USA, George Mason University, USA (2003)
3. Gacs, P., Kurdyumov, G., Levin, L.: One-dimensional uniform arrays that wash out finite islands. In: Problemy Peredachi Informatsii 12. (1978) 92–98
4. Goebels, A., Büning, H.K., Priesterjahn, S., Weimer, A.: Towards online partitioning of agent sets based on local information. In: Proceedings of the International Conference on Parallel and Distributed Computing and Networks (PDCN), Innsbruck, Austria (2005)
5. Mitchell, M., Hraber, P.T., Crutchfield, J.P.: Revisiting the edge of chaos: Evolving cellular automata to perform computations. (1993)
6. Boris, J.: A vectorized 'near neighbors' algorithm of order n using a monotonic logical grid. J. Comput. Phys. **66** (1986) 1–20
7. Andre, D., III, F.B., Koza, J.: Discovery by genetic programming of a cellular automata rule that is better than any known rule for the majority classification problem. In: In Proceedings of the First Annual Conference on Genetic Programming, MIT Press (1996)
8. Juille, H., Pollack, J.: Coevolving the ideal trainer: Application to the discovery of cellular automata rules. In: Proceedings of the Third Annual Genetic Programming Conference (GP-98). (1998)
9. Pagie, L., Mitchell, M.: A comparison of evolutionary and coevolutionary search. In: R.K. Belew and H. Juillé (editors): Coevolution: Turning Adaptive Algorithms upon Themselves, USA (2001) 20–25
10. Werfel, J., Mitchell, M., Crutchfield, J.: Resource sharing and coevolution in evolving cellular automata. In: IEEE Transactions on Evolutionary Computation, 4(4):388. (2000)
11. Kirchner, M.: Genetisch gesteuertes Lernen von Regeln für eindimensionale zelluläre Automaten. Bachelor thesis, Universität Paderborn (2005)
12. Bonabeau, E., Dorigo, M., Theraulaz, G.: Swarm Intelligence - From natural to artificial Systems. Oxford University Press (1999)
13. Kennedy, J., Eberhart, R.C.: Swarm Intelligence. Morgan Kaufmann (2001)
14. Wang, Y., Li, X.Y.: Distributed spanner with bounded degree for wireless ad hoc networks. In: Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing. (2002)
15. Eppstein, D.: Spanning trees and spanners. Technical Report Technical Report 96-16, University of California, Dept. Information and Computer Science (1996)

A Novel Clonal Selection for Multi-modal Function Optimization

Meng Hong-yun[1], Zhang Xiao-hua[2], and Liu San-yang[1]

[1] Dept.of Applied Math. Xidian University, Xian, China
mhyxdmath@hotmail.com
[2] Institute of Intelligent Information Processing, Xidian University, Xian, China
xh_zhang@mail.xidian.edu.cn

Abstract. This paper proposes a Clonal Selection Algorithm for Multimodal function optimization (CSAM) based on the concepts of artificial immune system and antibody clonal selection theory. In CSAM, more attention is paid to locate all the peaks (both global and local ones) of multimodal optimization problems. To achieve this purpose, new clonal selection operator is put forward, dynamic population size and clustering radius are also used not only to locate all the peaks as many as possible, but assure no resource wasting, i.e., only one antibody will locate in each peak. Finally, new performances are also presented for multimodal function when there is no prior idea about it in advance. Our experiments demonstrated that CSAM is very effective in dealing with multimodal optimization regardless of global or local peaks.

1 Introduction

At present, many researchers have gone into this field, and proposed some effective methods to solve multimodal function, such as crowding factor model proposed by De Jong [1], shared function model proposed by Goldberg and Richardson [2], artificial immune system method [3,4], split ring parallel genetic algorithms [5], Evolution Strategy [6], Particle swarm Optimization [7,8], and so on, which are trying to keep the diversity of population in the process of evolution. All above methods are effective in keeping diversity of population, and help a lot to locate more than one global peak. However, they seldom consider finding the sub-optima if a problem has any, while in the practical engineering, the sub-optimal solution maybe the alternative one. Thus, the main focus of this paper is to find a feasible method to locate all peaks of multimodal function no matter whether these peaks have the same height or not.

This paper is organized as follows. After defining several basic concepts of artificial immune system in Section 2, a novel algorithm, Clonal Selection Algorithm for Multimodal function optimization (CSAM), is put forward in Section 3. In Section 4, new performances are proposed, then three representative multimodal problems are selected for simulation tests in Section5. Finally, some discussions are presented in Section6.

2 Basic Definitions

The antibody clonal selection theory (F. M. Burnet, 1959) is used in the immune system to describe the basic features of an immune response. Its main idea lies in that the

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 63 – 72, 2006.

antigens can selectively react to the antibodies, which are the native production and spread on the cell surface in the form of peptides. The reaction leads to cell proliferating clonally and the clone has the same antibodies. Some clonal cells divide into antibodies that produce cells, and others become immune memory cells to boost the second immune response. In order to describe the algorithm well, some definitions are given in the following.

Definition1. Antigen. In Artificial immune system (AIS), antigen usually means the problem and its constraints. Without loss of generality, we consider maximizing multimodal optimization problems as follows.

$$\max_{x\in S} f(x), x = (x_1, x_2, \ldots, x_n) \tag{1}$$

where S is searching space. The antigen is defined as objective function $f(x)$, namely, $G(x) = g(f(x))$. Similar to the function of antigen in immunology, it is the initial factor for the artificial immune system algorithm. Usually, we let $G(x) = f(x)$ when not mentioned especially.

Definition2. **Antibody**. Antibodies represent candidates of the problem. The limited length character string $p = q_1 q_2 \ldots q_l$ is the antibody coding of variable x, denoted by $p = en(x)$, and x is called the decoding of antibody p, expressed as $x = de(p)$. In practice, binary coding and decimal coding are often used. Set I is called antibody space and the antibody population $PoP = \{p_1, p_2, \ldots, p_k\} \in I^k$ is a k-dimensional group of antibody p, namely

$$I^k = \{P : P = (p_1, p_2, \ldots, p_k), p_l \in I, 1 \le l \le k\} \tag{2}$$

where the positive integer k is the antibody population size.

Definition3. Antibody-Antigen Affinity. Antibody-Antigen Affinity is the reflection of the total combination power locates between antigen and antibodies. In AIS, it generally relates to the objective value or fitness. Namely

$$\text{Affinity}_{\text{Antibody-Antigen}}(p) = G(f(x)) \tag{3}$$

Definition4. Antibody-Antibody Affinity. Antibody-Antibody Affinity is the reflection of the total combination power locates between two antibodies. In this paper, we calculate the antibody-antibody affinity with Euclidean distance between $p = q_1 q_2 \ldots q_l$ and $P' = q_1' q_2' \ldots q_l'$, namely,

$$Affinity(p, p') = dist(p, p') = dist(de(p), de(p')) = \sqrt{\sum_{i=1}^{n} (x_i - x_i')^2} \tag{4}$$

Of course, other distance measures are possible.

3 Algorithm Description

3.1 Clustering Algorithm

As we known, there exist difficulties in searching all peaks for multimodal problem. One is how to keep the diversity of population. Another is how to avoid the less peak points being swallowed by the bigger one and the other is the number of peak points is unknown. In this paper a clustering algorithm is developed.

Let population in the $t-th$ generation as $PoP^t = \{p_1, p_2, \ldots, p_k\} \in I^k$, and num be the number of clustering. The main steps for it are in the following.

Step1. $num = 0, C = PoP^t = \{p_1, p_2, \ldots, p_k\}$.

Step2. $num = num + 1$. Find the best antibody P^{num} in the current antibody set C, satisfying $G(f(de(p^{num}))) = \max_{p \in C} G(f(de(p)))$ and construct Clustering set C_{num} as follows,

$C_{num} = \{p \mid Affinity(p, p^{num}) < \omega, p \in C\}$ $C_{Temp} = \{p \in C, and, p \notin C_{num}\}, C = C_{Temp}$

Stp3. If $C \neq \phi$, go to step2; else stop.

Based on the above operations, num clustering set C_i ($i = 1, \ldots, num$) centered as P^i can be obtained, and $\bigcup_{i=1}^{num} C_i = PoP^t, C_i \bigcap C_j = \phi, i \neq j$, ω is clustering radius.

3.2 The Framework of CSAM

Inspired from the biology immune system and the clonal selection mechanism, CSAM is put forward in this section. The main process for Clonal Selection Operator is presented as Fig. 1, and the antibody population at time k is represented by the time-dependent variable $P(k) = \{P_1(k), P_2(k), \ldots, P_n(k)\} = \bigcup_{i=1}^{num} C_i$.

Clonal Operating T_c^C : Define

$$T_c^C(P(k)) = [T_c^C(P_1(k)) \quad T_c^C(P_2(k)) \quad \cdots \quad T_c^C(P_n(k))]^T \tag{5}$$

where $T_c^C(P_i(k)) = I_i \times P_i(k) (i = 1,2 \cdots, n)$, I_i is a q_i dimension row vector. $T_c^C(P_i(k))$ is the q_i clone of antibody and q_i is adjusted to the antibody-antigen affinity adaptively. After Clonal Operating, the population becomes $\overline{P}_{new}(k) = \{\overline{P}_1(k), \overline{P}_2(k), \cdots, \overline{P}_n(k)\}$ where $\overline{P}_i(k) = \{P_{i1}(k), P_{i2}(k), \cdots, P_{iq_i}(k)\}$ and $P_{ij} = P_i(k), j = 1,2, \cdots, q_i$.

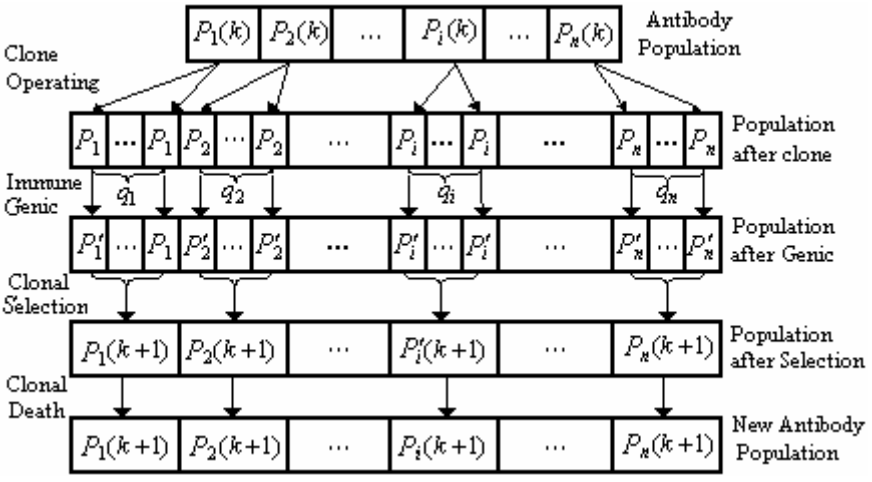


Fig. 1. The main operating processing of the Clonal Selection Operator

Immune Genic Operating T_g^C : Crossover and Mutation are the two main operators in immune genic operating. In this paper, all the antibodies can be categorized into three groups, and different operators are adopted on the different groups. The combination of the three operators helps to improve the global search ability.

For antibody set $C_i = \{p \mid Affinity(p, p^i) < \omega, p \in P(k)\}$ centered on p^i ($i = 1, \ldots, num$), let $x = de(p^i)$ and for every antibody $p \in C_i$, if $p = p^i$, we adopt Gaussian mutation to generate new antibody $p' = en(x')$ in the following

$$x' = x + \omega * g(0, \sigma) \tag{6}$$

where $g(0, \sigma)$ is a Gaussian random variable.

To antibodies in $\{p \mid Affinity(p, p_i) < \omega / 2, p \neq p^i, p \in C_i\}$, study operator is adopted as follows

$$x' = x_m + c * r(x - x_m) \tag{7}$$

Here, r is a uniformly random value in the range [0, 1]. c is study factor.

And to antibodies in set $\{p \mid affinity(p, p_i) \geq \omega / 2, p \in C_i\}$, uniformly mutation is adopted according to the following equation.

$$x' = x_{low} + r(x_{up} - x_{low}) \tag{8}$$

x_{up}, x_{low} is upper bound and lower bound of x, r is a uniformly random value between 0 and 1.

Clonal Selection Operating T_s^C **:** $\forall i = 1,2,\cdots,n$, if there exists mutated antibody B_i which satisfies $f(de(B_i)) = \max f(de(P_{ij}))$ $j = 1,2,\cdots,q_i$ .The probability of B_i taking place of $P_i(k)$ is defined as

$$p(P_i(k) \to B) = \begin{cases} 1 & f(de(P_i(k))) \le f(de(B_i)) \\ \exp\left(\dfrac{f(de(P_i(k))) - f(de(B_i))}{a}\right) & f(de(P_i(k))) > f(de(B_i)) \end{cases} \tag{9}$$

Here α is a positive value related to the diversity of the antibody population. Generally, the better the diversity is, the bigger α is. Otherwise, α is smaller.

Clonal Death Operating T_d^C **:** After the above operating, antibody population is $\overline{P}(k+1) = \{p_1(k+1), \cdots, p_n(k+1)\}$. If $dist(P_i(k+1), P_j(k+1)) < \varepsilon$, one of $P_i(k+1)$ and $P_j(k+1)$ will be died forcedly according to the death probability p_d , where $P_i(k+1)$ and $P_j(k+1)$ belong to $P(k+1)$, ($i \neq j, i, j = 1,2,\cdots,n$), $\varepsilon > 0$. The death strategy can be either generating a new antibody randomly to replace $P_i(k+1)$ or $p_j(k+1)$, or using mutation strategy to generate a new antibody to replace them, and thus we complete the whole operation on antibody population.

3.3 Self-adaptive for ω and Population Size

Another difficulty for multimodal problem is that the number of extreme points is unknown to us, which requires the scale of population to be adjusted adaptively. Generally, there are two cases. On the one hand, if the population size is too big, computation resources may be wasted. On the other hand, if too small, some peaks may be ignored. Therefore, an adaptive operator will be adopted for the scale of the population, and clustering radius will also be adjusted adaptively.

3.3.1 Self-adaptive for Population Size

There is an important question: how many new individuals should be introduced? In order to guarantee locating all peaks and wasting no resources, a dynamic strategy will be taken in this paper.

In this context, if only P^i is in the clustering set C_i , then the size of the population will be increased with the following equation

$$PoPSize(k+1) = PoPSize(k) + \mu * PoPSize(k) \tag{10}$$

where $PoPSize(k)$ is the size in generation iteration k , μ is a learning factor which is related to the number of clustering set, and the new individuals were generated randomly to improve the ability of global searching .

On the other hand, when the number for clustering set is far less than the population size and its number is no longer increase during many generations, then the population should be decreased according to the following

$$PoPSize(k+1) = PoPSize(k) - \mu * PoPSize(k) \tag{11}$$

Generally, individuals approaching to the clustering center will be deleted.

3.3.2 Self-adaptive for ω

If the population size is up to the predefined value $num_{\max}$, the clustering radius need be decreased, here $x_i \in [low_i, up_i]$, and $num_{\max}$ is defined as follows

$$num_{\max} = \prod_{i=1}^{n} \left(\frac{up_i - low_i}{2\omega} \right) \tag{12}$$

Obviously, a fixed ω is not suitable for searching more extreme points, and too large clustering radius ω may lead some extreme points to be swallowed up. While smaller ω makes convergence slowly. Hence, new method is taken in our paper as follows. Let $P = \{p^1, \ldots, p^{num}\}$ be the set of all clustering center in a certain generation, if one clustering has no improvement during the generation, then clustering radius should be reduced, otherwise, it will be increased.

The total process of the proposed CSAM is in the following:

Step1 t=0, initialize antibody population $PoP^t = \{p_1, p_2, \ldots, p_k\} \in I^k$, compute the affinity among all antibodies and set all relative parameters;

Step2 use Clustering Algorithm, the population is divided into num sub-population $PoP^t = \bigcup_{i=1}^{num} C_i$, the Clustering centre is p^i.

Step3 According to affinities and given clonal size, implement Clonal Operating T_c^C, Immune Genic Operating T_g^C, Clonal Selection Operating T_s^C, Clonal Death Operating T_d^C, thus obtains the new antibody population is denoted as $P(k)$.

Step4 Computing the affinity of $P(k)$, Self-adaptive updating of Population Size and Clustering radius.

Step5 k=k+1；If the halt conditions achieved, then stop, otherwise go to step2.

4 Performance Measurements

Generally, the performance for multimodal function can be measured according to three criteria, number of evaluations required to locate the peak, accuracy, measuring the closeness to the peak, and success rate, i.e., the percentage of runs in which all global optima are successfully located.

In this paper, new performances for measuring the closeness to the optima are presented according to two cases. One is all the extreme points are all known in advance and the other is that there is no any information to the extreme points in advance.

In the first case, assume $P_{extremum} = \{p_1^e, \ldots p_N^e\}$ includes all the extreme points, N is the number of extreme points, and the approximation set of the extreme points is $P_{proposed} = \{p_p^1, \ldots p_p^M\}$, M is the cardinality of $P_{proposed}$, then the approximation of $P_{proposed}$ to $P_{extremum}$ can be defined as follows

$$GD = \sum_{i=1}^{M} dist(p_p^i, P_{extremum}) / M \tag{13}$$

where $dist(p_p^i, P_{extremum}) = \min_{q \in P_{extremum}} d(p_p^i, q)$. $d(p_p^i, q)$ is the distance between two antibody p_p^i and q.

At the same time, the success rate is defined as

$$P_{rate} = |P_{find}| / N \tag{14}$$

Here $P_{find} = \bigcup_{i=1}^{M} \{p \mid d(p, p_p^i) < d(q, p_p^i), p, q \in P_{extremum}, p_p^i \in P_{proposed}\}$.

In the second case, let $P_{proposed} = \{p_p^1, \cdots, p_p^M\}$. Obviously, if the p_p^i is approaching some real extreme point, the bigger probability of $\{f(de(p_p^i)) > f(de(q))\}$ is, here q is a random antibody in the neighborhood centered with p_p^i.Obviously, the probability equal to one means the p_p^i is a real peak.

Based on the above consideration, N points are selected randomly from the neighborhood of p_p^i with certain distribution (Gaussian or uniform•, which are denoted as $P_{rand} = \{p_{rand}^1, \ldots p_{rand}^N\}$.

Computing the probability

$P_r^i = \sum_{i=1}^{M} \frac{|\{p \mid f(de(p_p^i)) > f(de(p)), p \in P_{rand}\}|}{|P_{rand}|}$ where $p_p^i \in P_{proposed}$, $|P_{rand}|$ is the cardinality of P_{rand}. The approximation of set $P_{proposed}$ can be measured with

$$P_r = \sum_{i=1}^{M} p_r^i / M \tag{15}$$

5 Numerical Tests and Results

To validate our approach, we chose three multimodal functions, which turn out to be difficult for any search algorithm because they have numerous peaks. For convenience, decimal coding is used and the related parameters are in the following: population size is 100 and function evaluation times are 25201, clustering radius $\omega = 0.1$, learning factor $\mu = 0.1$, death probability p_d is 0.8 . And mutation is according to the

equation $x = x + 0.1 * gaussian(0,0.1)$, study operator is according to $x = x + w * 2 * rand(0,1) * (p - x)$. In addition, we found an interesting case during the experiments, when ω is chosen as 0.4 , study operator is similar to particle swarm optimization.

5.1 Discussion of the Results

In this paper, three functions are used in the following, the first one is used in [4].the second is Rastrigin function and the last one is Schwefel function.

(1) $F(x, y) = x\sin(4\pi x) - y\sin(4\pi y) + 1$, where $x, y \in [-2,2]$.

(2) $G(x, y) = 20 + x^2 - 10\cos 2\pi x + y^2 - 10\cos 2\pi y$, where $x, y \in [-5.12,5.12]$.

(3) $H(x, y) = -\left(x\sin\sqrt{|x|} + y\sin\sqrt{|y|}\right)$, where $-500 \le x, y \le 500$.

The first function was taken frequently by many researchers, which has many local optima peaks and one global optimum peak. Fig. 2(a) illustrates the located peaks of the proposed algorithm for function $F(x, y)$. Fig. 2(b) gives the location of found peaks in variation space. In [5], only 18 peaks are found, and in [4], 61peaks were located, while in this paper, 80 peaks are located.

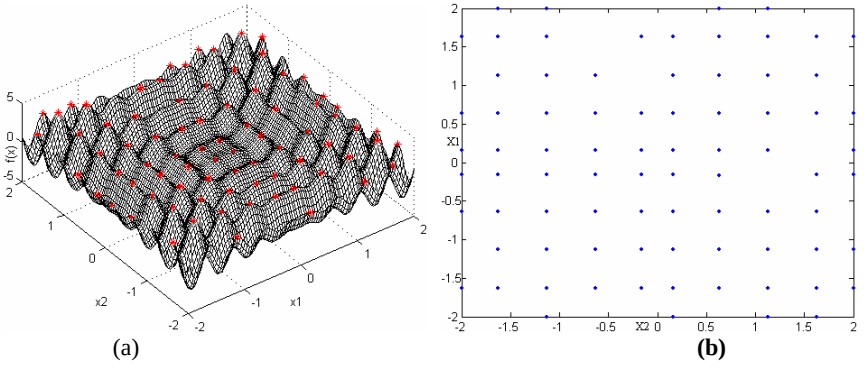


Fig. 2. Simulation of function F(x, y) (a) Plots of function F(x, y) (b) distribution of antibodies in variation space

The second function has a lot of local and global peaks and the third not only has many local peaks, but also the range of variation is wide, due to much difficult to optimize, few report is given. Fig.3(a), 4(a) illustrates the located peaks found by the proposed algorithm for them and Fig.3(b), 4(b) gives the location of found peaks in variation space. From Fig. 2(a), 3(a) and 4(a), it can be seen that our algorithm could find almost all peaks.

By using Self-adaptive adjustment of Population Size, the number of peaks found is almost equal to the size of population. To the first function, 80 peaks are located, the Population Size is 90; to the second function 96 peaks are located, the Population Size is 100; to the last function, 54 peaks is found, and the Population size is 60.

In order to illustrate the feasibility and efficiency of our algorithm,Table1 show the min, max, mean, median and standard deviation (std.dev) for the above functions, regarding to the approximation of our method, for the first function, the mean can achieve 0.9826, the standard deviation is 0.0366. For the second function, the mean is 0.9993, the standard deviation is 0.0015. To the third function, the mean can achieve 0.9883, the standard deviation is 0.0284.

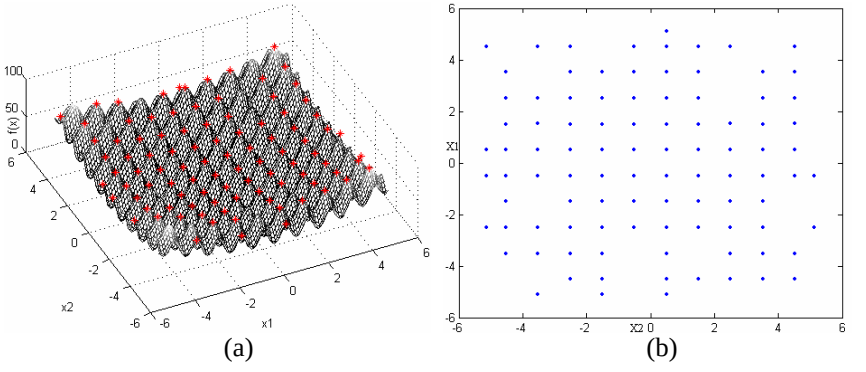


Fig. 3. Simulation of Rastrigin function. (a) plots of Rastrigin function (b) distribution of antibodies in variation space.

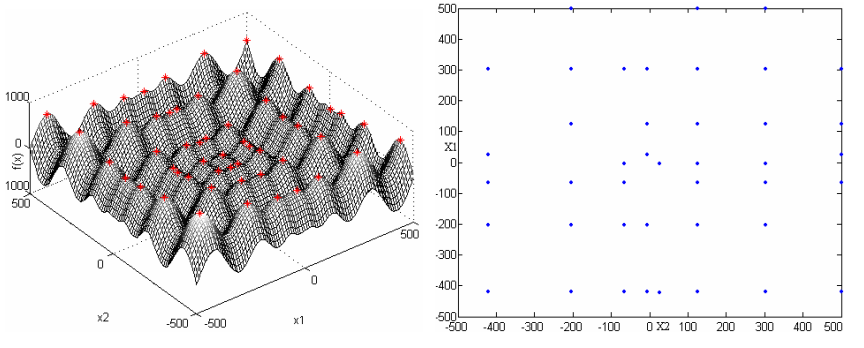


Fig. 4. Simulation of Schwefel function. (a) Plots of Schwefel function (b) distribution of antibodies in variation space.

Table 1. Approximation of the proposed algorithm for different functions

Function	Number of peaks found	Approximation of the proposed algorithm				
$F(x, y)$	80	0.8530	1.0000	0.9826	1.0000	0.0366
$G(x, y)$	96	0.9930	1.0000	0.9993	1.0000	0.0015
$H(x, y)$	54	0.8940	1.0000	0.9883	1.0000	0.0284

6 Conclusion

This paper presents a novel algorithm for multimodal function optimization. Based on the AIS, a novel clonal selection are designed, in addition, to keep the diversity of population and avoid extreme points with different fitness values to be swallowed up, clustering operators and population size are adjusted adaptively. To measure the performance of the different algorithm for multimodal problem, new performances were also proposed. Simulations show that our algorithm is competitive, and it can find peaks as many as possible with the advantage of easy to operate and few parameters to adjust. However, how to judge a point is a globally one, how to select a suitable clustering radius and how to measure the convergence are our next work.

References

1. De Jong, K.A., An Analysis of the Behavior of a Class of Genetic Adaptive Systems. Ph.D. Thesis, Ann Arbor, MI: University of Michigan. Dissertation Abstracts International 36(10), 5410B(University Microfilms No. 76-9381)
2. Goldberg, D.E. and Richardson, J., Genetic Algorithms with Sharing for Multimodal Function Optimization. In Proceedings of the Second International Conference on Genetic Algorithms and Their Applications, 1987, 41-49.
3. Toyoo Fukuda, Kazuyuki Mori and Makoto Tsukiyama, Parallel Search for Multi-Modal Function Optimization with Diversity and Learning of Immune Algorithm, Artificial Immune Systems and Their Applications, 210~220, D.DASGUPTA Ed., Springer-Verlag Berlin Heidelberg, 1999.
4. Leandro N. de Castro and Jon Timmis, An Artificial Immune Network for Multimodal Function Optimization, Russ Eberhart Ed., Proceedings of the 2002 Congress on Evolutionary Computation, May 12-17, 2002, Honolulu, Hawaii, IEEE Service Center, 2002, 699-704.
5. K.Burton Havrvey & Chrisila C. Pettey, The Outlaw Method for Solving Multimodal Function with Split Ring Parallel Genetic Algorithms, in Proceedings of GECCO' 99(the Genetic and Evolutionary Computation Conference), July 13-17, 1999, Orlando, Florida, Morgan Kaufmann Publishers, 274-288.
6. Chang-Hwan Im, Hong-Kyu Kim, Hyun-Kyo Jung, and Kyung Choi, A Novel Algorithm for Multimodal Function Optimization Based on Evolution Strategy, IEEE Trans on magnetics, vol. 40, no. 2, march 2004,1224-1227.
7. Higashi N., Iba H., Particle Swarm Optimization with Gaussian Mutation. In IEEE Swarm Intelligence Symposium (2003) 72-79.
8. Susana C.E., Carlos A. C. C., On the Use of Particle Swarm Optimization with Multimodal Functions. Proceedings of IEEE Congress on Evolutionary Computation 2003 (CEC 2003), Canbella, Australia. pp. 1130-1136, 2003.
9. De Castro, L.N., & Von Zuben, F. J., Learning and Optimization Using the Clonal Selection Principle. IEEE Trans. On Evol. Comp., Special Issue on Artificial Immune System, 6(3):239-251, 2001.

Grid Intrusion Detection Based on Immune Agent

Xun Gong, Tao Li, Tiefang Wang, Jin Yang, Gang Liang, and Xiaoqin Hu

School of Computer Science, Sichuan Univ., Chengdu 610065, China
orchid_xun@tom.com

Abstract. This paper proposes a novel grid intrusion detection model based on immune agent (*GIDIA*), and gives the concepts and formal definitions of *self*, *nonself*, *antibody*, *antigen*, *agent* and *match algorithm* in the grid security domain. Then, the mathematical models of *mature MoA* (mature monitoring agent), and *dynamic memory MoA* (memory monitoring agent) *survival* are established. Besides, effects of the important parameter T in the model of mature MoA on system performance are showed. Our theoretical analysis and experimental results show that the model that has higher detection efficiency and steadier detection performance than the current models is a good solution to grid intrusion detection.

1 Introduction

Grid environments have many distinctive characteristics that are different from common network environments [1], so grid intrusion detection problems that are not addressed by existing intrusion detection technologies for common network environments. Artificial immune system has the features of dynamicity, self-adaptation and diversity [2-7] that just meet the constraints derived from the characteristics of the grid environment, and mobile agent has many same appealing properties as that of artificial immune system [8]. Thus, we apply mobile agent technology as support for intrusion detection, and propose a novel grid intrusion detection model based on immune agent (*GIDIA*).

In *GIDIA*, the concepts and formal definitions of *self*, *nonself*, *antibody*, *antigen* and *agent* (that simulates the lymphocyte and is used as a detector to recognize nonself antigens, i.e. intrusions,) in the grid security domain are given firstly. We define three kinds of immune agents: monitoring agents (*MoA*), communicator agents (*CoA*) and beating off agents (*BoA*). Then, the mathematical models of *mature MoA* and *dynamic memory MoA survival* are established.

In the present immunity based intrusion detection systems [9], the memory lymphocytes have an unlimited lifecycle except they match the newly added selfs. Obviously, a considerable number of memory lymphocytes will be generated in the end. As the increasing of memory cells, the cells number will increase, at the same time, the intrusion detection time will increase. When the number of cell reaches a certain value, the intrusion detection system will either become a bottleneck or ineffective as some packets are skipped.

However, *GIDIA* introduces a new idea of *dynamic memory MoAs survival* to overcome this problem: a given number of the least recently used memory MoAs that

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 73 – 82, 2006.

simulate memory lymphocytes will degrade into mature MoAs and be given a new age and affinity (the number of antigens matched by antibody), if the number of MoAs reaches or exceeds a given maximum. The method assures the number of memory cells does not exceed a given value, so that the total number of cells does not exceed a given value. So it does not take too long time to detect intrusions. The new age of degraded memory MoAs cannot be set too large or too small. Too large age will affect the diversity of *GIDIA* and too small age will reduce detection efficiency, which are proved in *Theorem 1*. In mature MoA model, the new affinity of degraded memory MoAs is set close to the lifecycle of mature MoAs. Once a degraded memory MoA matches only one antigen in its lifecycle, it will be activated immediately and evolve into memory one again. It is important to set the number of degraded memory MoAs to a given value, which assures *GIDIA* steady performance. If the number of degraded memory MoAs is not given, the number of degraded memory MoAs will be random and the remainder will be small or large. The false negative rate will be large or intrusion detection time will be long. The set of new affinity gives priority to degraded memory MoAs in evolvement, which enhances detection efficiency.

2 The Proposed Grid Intrusion Detection Model (*GIDIA*)

GIDIA has two sub-models (see Fig. 1). The first is the model of the generation and evolvement of MoAs in the square frame. The second is the intrusion detection model.

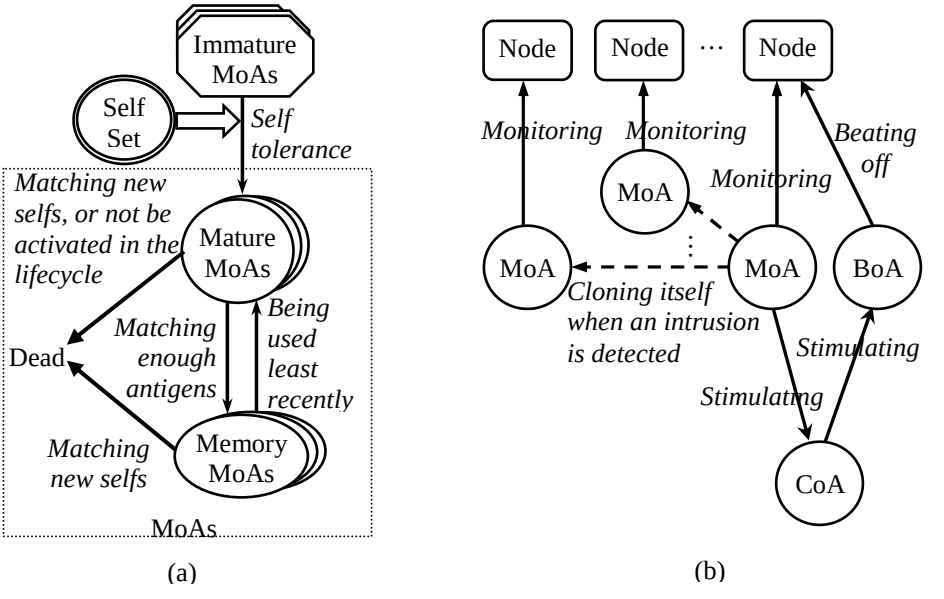


Fig. 1. The proposed grid intrusion detection model. (a) The generation and evolvement of MoAs. (b) Intrusion detection model.

In the second sub-model, we define antigens (*Ag*) to be the features of grid services and accesses:

$$Ag = \{ag \mid ag \in D\}. \tag{1}$$

$$D = \{d \mid d = < uslevel, syslevel, prolevel, paclevel >\}, D = \{0,1\}^l. \tag{2}$$

uslevel, *syslevel*, *prolevel* and *paclevel* present the parameters, which are monitored by monitoring agents at user, system, process, and packet level respectively. They are also called, respectively, the field of an antigen. *l* is a natural number (constant). The length of string *ag* is *l*. For the convenience using the fields of a antigen *x*, a subscript operator "." is used to extract a specified field of *x*, where

$$x.fieldname = \text{the value of field } fieldname \text{ of } x. \tag{3}$$

For example, *x.uslevel = 01101…10*.

We define *Self* to be the set of normal grid services and accesses. Similarly, *Nonself* is a set of abnormal services and accesses. *Ag* contains two subsets, *Self* and *Nonself*, where $Self \subset Ag$ and $Nonself \subset Ag$ such that

$$Self \cup Nonself = Ag, \quad Self \cap Nonself = \Phi. \tag{4}$$

In the second sub-model, MoAs simulate B-lymphocytes in the immune system and patrol grid nodes. They are responsible for monitoring various parameters simultaneously at four levels (user level, system level, process level, and packet level) and detecting intrusion. As B-lymphocytes consist of mature and memory lymphocytes, MoAs are divided into mature and memory MoAs that simulate, respectively, mature and memory cells. CoAs serve as communicators and are responsible for message transmission among agents. They simulate lymphokines secreted from T cells to stimulate B cells to clone themselves or eliminate cloned cells. BoAs simulate T killer cells are responsible for dealing with intrusive activities. Once MoAs detect intrusions, they will stimulate CoAs and present the features of intrusions to BoAs. CoAs will activate BoAs. BoAs will move to the suspected place to counterattack intrusions. The counterattack strategies consist of disconnecting a node, killing a process, discarding dubitable packets, and so on. MoAs can clone themselves if the intrusion is detected. Some cloned MoAs are left here to detect more intrusions, and others are moved to other grid nodes to detect the similar intrusions.

All the agents form a set (*Agent*). Agent contains three elements, *MoA*, *CoA* and *BoA*. Monitoring agents, communicator agents and beating off agents form, respectively, the set *MoA*, *CoA* and *BoA*. A monitoring agent is used as a detector to recognize nonself antigens (intrusions). Thus, we have:

$$Agent = \{MoA, CoA, BoA\}. \tag{5}$$

$$CoA = \{c \mid c \text{ is a communicator agent which transmits message among agents.}\}. \tag{6}$$

$$BoA = \{b \mid b \text{ is a beating off agent which counterattacks intrusions according to the antigen features presented by MoAs. }\} \quad (7)$$

$$MoA = \{< d, age, count > \mid d \in D, age \in N, count \in N\}. \quad (8)$$

d is the lymphocyte antibody that is used to match an antigen. Each monitoring agent carries several antibodies. *age* is the agent age, *count* (affinity) is the antigen number matched by antibody *d*, and *N* is the set of natural numbers. *d*, *age* and *count* are also called, respectively, field *d*, *age* and *count* of an agent. Similarly, the subscript operator "." is used to extract a specified field of an agent *x*. For example, *x.d.uslevel = 1101…101*, *x. age = 25*.

MoAs consist of Mature and Memory MoAs (see sect. 1). A mature MoA is a MoA that is tolerant to self but is not activated by antigens. A memory MoA evolves from a mature one that matches enough antigens in its lifecycle. Mature and memory MoAs form, respectively, the set MA_{MoA} and ME_{MoA}. Mature MoAs detect novel intrusions that have not previously been identified. Memory MoAs greatly enhance detection of previously seen intrusions. Therefore, we have:

$$MoA = MA_{MoA} \cup ME_{MoA}, \qquad MA_{MoA} \cap ME_{MoA} = \Phi. \quad (9)$$

$$MA_{MoA} = \left\{x \mid x \in MoA, \forall y \in Self \quad (< x.d, y > \notin Match \wedge x.count < \beta)\right\}. \quad (10)$$

$$ME_{MoA} = \left\{x \mid x \in MoA, \forall y \in Self \quad (< x.d, y > \notin Match \wedge x.count \geq \beta)\right\}. \quad (11)$$

Match is a match relation in *D* defined by

$$\begin{aligned} Match = \{< x, y > \mid (x, y \in D),\ & (f_{match}(x.uslevel, y.uslevel) = 1 \\ \vee\ & f_{match}(x.syslevel, y.syslevel) = 1 \vee f_{match}(x.prolevel, y.prolevel) = 1 \\ \vee\ & f_{match}(x.paclevel, y.paclevel) = 1)\}. \end{aligned} \quad (12)$$

$f_{match}(x, y)$ is based on the affinity between *x* and *y*: if the affinity greater than a specified threshold, then 1 is returned, otherwise, 0 is returned. In *GIDIA*, the affinity function can be *r-contiguous-bits matching rule*, *Hamming distance*, *Landscape-Affinity Matching*, etc. As intrusions do not always happen simultaneously at four levels, we consider an intrusion happens as long as MoA matches an antigen at any one level of the fours. The method enhances detection efficiency.

2.1 Mature MoAs Model

$$MA_{MoA}(t) = \begin{cases} \Phi, & t = 0 \\ MA_{retain}(t) \cup MA_{new}(t) \cup MA_{cycle}(t) - MA_{activation}(t), & t \geq 1 \end{cases}. \quad (13)$$

$$MA_{retain}(t) = MA'_{MoA}(t) - A(t) \cup A'(t). \quad (14)$$

$$MA'_{MoA}(t) = \{y \mid y \in MoA, x \in MA_{MoA}(t-1), x.age < \lambda, \forall s \in Self_{new}(t) < x.d, s > \notin Match, \\ y.d = x.d,\ y.age = x.age + 1,\ y.count = x.count\}. \tag{15}$$

$$A(t) = \{x \mid x \in MA'_{MoA}(t), \exists y \in Ag(t-1)\ < x.d, y > \in Match\}. \tag{16}$$

$$A'(t) = \{y \mid y \in MoA, x \in A(t), y.d = x.d,\ y.age = x.age, y.count = x.count + 1\}. \tag{17}$$

$$MA_{cycle}(t) = \left\{y \middle| y \in MoA, y.d = x.d, y.age = \mathrm{T}, y.count = \beta - 1, x \in ME_{degradation}(t), \mathrm{T} < \lambda\right\}. \tag{18}$$

$$MA_{activation}(t) = \{x \mid x \in A'(t),\ \ x.count \geq \beta\}. \tag{19}$$

$$MA_{new}(t) = \{y \mid y \in MoA, y.d = x.d, y.count = 0, y.age = 0, \\ x \in (I_{new}(t) - \{z \middle| z \in I_{new}(t), \exists s \in Self(t) < z, s > \in Match\})\}. \tag{20}$$

$$I_{new}(t) = \begin{cases} \Phi, & |MA_{MoA}(t-1)| \geq \Gamma \\ \{y_1, y_2, ..., y_{\Gamma - |MA_{MoA}(t-1)|}\}, y_i \in D,\ 1 \leq i \leq \Gamma - |MA_{MoA}(t-1)|, & otherwise \end{cases}. \tag{21}$$

$MA_{retain}(t)$ simulates the process that the mature cells evolve into the next generation ones, where the cells do not tolerate to those newly added self elements or have not match enough antigens ($\beta > 0$) in lifecycle λ , will be eliminated. $MA_{new}(t)$ depicts the generation of new mature MoAs. Γ (>0) is the max number of mature MoAs in *GIDIA*. $MA_{activation}(t)$ is the set of mature MoAs which match enough antigens and will be activated and evolved into memory MoAs at time t. $MA_{cycle}(t)$ is the set of the least recently used memory MoAs which degrade into mature MoAs and be given a new age T (>0) and count $\beta - 1$ ($\beta \geq 1$). Once a mature MoA with count $\beta - 1$ matches an antigen in lifecycle λ , the count of the mature MoA will be set to β and it will be activated again. Because the degraded memory MoA has better detection capability than mature MoAs, it is endowed with a priority in evolvement into memory MoAs. When the same antigens arrive again, they will be detected immediately by the re-evolved memory MoA. The method enhances detection efficiency.

Theorem 1. Given all the degraded memory MoAs have tolerated to those newly added self elements from time t_1 to t_n , T is in direct proportion to $|MA_{new}(t_n + 1)|$.

Proof: Suppose the least recently used memory MoAs degrades into mature MoAs at the time point t_1, t_2, …, t_n, respectively. Thus, according to equations (13) - (21), we have:

$$MA_{cycle}(t) = \Phi \quad (t \in [t_1 + 1, t_n - 1] \wedge t \neq t_2, t_3, \ldots, t_{n-1}). \tag{22}$$

$$MA_{retain}(t) = MA_{cycle\_tolerance}(t) \cup MA_{other\_retain}(t) \\ (t \in [t_1 + 1, t_n - 1] \wedge t \neq t_2, t_3, \ldots, t_{n-1}). \tag{23}$$

$$MA_{MoA}(t) = MA_{cycle\_tolerance}(t) \cup MA_{other}(t) \quad (t \in [t_1 + 1, t_n - 1] \wedge t \neq t_2, t_3, \ldots, t_{n-1}). \tag{24}$$

$MA_{cycle\text{-}tolerance}(t)$ is the set of memory MoAs which are degraded at the time point *t-1* and tolerate to those newly added self elements at the time point *t*. $MA_{other\text{-}retain}(t)$ is the set of other retained mature MoAs. $MA_{other}(t)$ is the set of other mature MoAs at the time point *t*. At the time point t_1, we have:

$$MA_{cycle}(t_1) \neq \Phi. \tag{25}$$

At the time point t_2, we have:

$$MA_{cycle}(t_2) \neq \Phi. \tag{26}$$

$$MA_{retain}(t_2) = MA_{cycle\_tolerance}(t_2) \cup MA_{other\_retain}(t_2). \tag{27}$$

$$MA_{MoA}(t_2) = MA_{cycle\_tolerance}(t_2) \cup MA_{cycle}(t_2) \cup MA_{other}(t_2). \tag{28}$$

Similarly, at the time point t_3, we have:

$$MA_{cycle}(t_3) \neq \Phi. \tag{29}$$

$$MA_{retain}(t_3) = MA_{cycle\_tolerance}(t_3) \cup MA_{other\_retain}(t_3). \tag{30}$$

$$MA_{MoA}(t_3) = MA_{cycle\_tolerance}(t_3) \cup MA_{cycle}(t_3) \cup MA_{other}(t_3). \tag{31}$$

Suppose all the degraded memory MoAs have tolerated to those newly added self elements and been in mature MoAs from time t_1 to t_n in the worst condition. The survival time of the degraded memory MoAs is $t_n - t_1$. Therefore, we have:

$$t_n - t_1 = \lambda - \mathrm{T}. \tag{32}$$

And, we have:

$$MA_{cycle}(t_1) = MA_{cycle\_tolerance}(t_2). \tag{33}$$

$$MA_{MoA}(t_2) = MA_{cycle}(t_1) \cup MA_{cycle}(t_2) \cup MA_{other}(t_2). \tag{34}$$

$$MA_{cycle}(t_1) \cup MA_{cycle}(t_2) = MA_{cycle\_tolerance}(t_3). \tag{35}$$

$$MA_{MoA}(t_3) = MA_{cycle}(t_1) \cup MA_{cycle}(t_2) \cup MA_{cycle}(t_3) \cup MA_{other}(t_3). \tag{36}$$

Then, we have:

$$MA_{MoA}(t_3) = \bigcup_{i=1}^{3} MA_{cycle}(t_i) \cup MA_{other}(t_3). \tag{37}$$

Similarly, at the time point t_n, we have:

$$MA_{MoA}(t_n) = \bigcup_{i=1}^{n} MA_{cycle}(t_i) \cup MA_{other}(t_n). \tag{38}$$

$$\left|MA_{MoA}(t_n)\right| = \left|\bigcup_{i=1}^{n} MA_{cycle}(t_i)\right| + \left|MA_{other}(t_n)\right|. \tag{39}$$

Obviously, T is in inverse proportion to *n* according to equation (32). And *n* is in direct proportion to $\left|\bigcup_{i=1}^{n} MA_{cycle}(t_i)\right|$. $\left|\bigcup_{i=1}^{n} MA_{cycle}(t_i)\right|$ is in direct proportion to $\left|MA_{MoA}(t_n)\right|$ according to equation (39). $|MA_{MoA}(t_n)|$ is in inverse proportion to the number $\left|I_{new}(t_n+1)\right|$ of new immature MoAs according to equation (21), and in inverse proportion to $\left|MA_{new}(t_n+1)\right|$ the number of new mature MoAs according to equation (20). So $\left|\bigcup_{i=1}^{n} MA_{cycle}(t_i)\right|$ is in inverse proportion to $\left|MA_{new}(t_n+1)\right|$. Similarly, T is in direct proportion to $\left|MA_{new}(t_n+1)\right|$.

The value of T cannot be set too large or too small. If the value of T is too large and even close to λ, the survival time of the degraded memory MoAs will be short, and the degraded memory MoA will die soon because it cannot match enough antigens in lifecycle λ. Contrarily, if the value of T is too small and even close to 0, the degraded memory MoA will have been in mature MoAs for a long time. As the degraded memory MoAs in mature MoAs increasing, the number of mature MoAs $\left|MA_{MoA}(t_n)\right|$ will increase. The larger $\left|MA_{MoA}(t_n)\right|$ is, the less new immature MoAs $I_{new}(t_n+1)$ and even new mature MoAs $MA_{new}(t_n+1)$ will be generated. When the number of mature MoAs reaches or exceeds the maximum Γ, immature MoAs will not be generated. Therefore, too large value of T will affect the diversity of *GIDIA*.

2.2 Dynamic Memory MoAs Survival Model

$$ME_{MoA}(t) = \begin{cases} \Phi, & t = 0 \\ ME_{retain}(t) \cup ME_{new}(t) - ME_{\deg radation}(t), & t \geq 1 \end{cases}. \tag{40}$$

$$ME_{retain}(t) = ME'_{MoA}(t) - E(t) \cup E'(t). \tag{41}$$

$$ME'_{MoA}(t) = \{y \mid y \in MoA, x \in ME_{MoA}(t-1), \forall s \in Self_{new}(t) < x.d.s > \notin Match, y.d = x.d, y.age = x.age + 1, y.count = x.count\}. \tag{42}$$

$$E(t) = \{x \mid x \in ME'_{MoA}(t), \exists y \in Ag(t-1) < x.d, y > \in Match\}. \tag{43}$$

$$E'(t) = \{ y \mid y \in MoA, x \in E(t), y.d = x.d, y.age = 0, y.count = x.count + 1 \}. \quad (44)$$

$$\begin{aligned} ME_{new}(t) = \{ x \mid x \in ME_{MoA}, y \in MA_{activation}(t), \\ x.d = y.d, x.age = 0, x.count = y.count \}. \end{aligned} \quad (45)$$

$$ME_{degradation}(t) = \begin{cases} \{x_1, x_2, \cdots, x_P\}, 1 \le i \le P, x_i \in ME_{degradation}(t), \\ \forall y \in ME_{retain}(t) - ME_{degradation}(t) y.age \le x_i.age, & |ME_{MoA}(t-1)| \ge \kappa \\ \Phi, & otherwise \end{cases}. \quad (46)$$

Equation (40) depicts the dynamic evolvement of memory MoAs, where the memory MoAs that match the newly added selfs will die and those matched by an antigen will be activated immediately. $ME_{retain}(t)$ simulates the process that the memory MoAs evolve into the next generation ones. $ME_{new}(t)$ is the set of new memory MoAs evolved from mature ones. $ME_{degradation}(t)$ is the set of memory MoAs that are not activated by antigens lately and are selected randomly to degrade into mature MoAs when the number of memory MoAs reaches a given maximum κ. P (>0) is the number of selected memory MoAs. Obviously, the number of memory MoAs does not exceed a certain maximum.

3 Simulations and Experiment Results

To prove the intrusion detection performance of *GIDIA*, we developed abundant experiments to show the effect of parameter T on high detection efficiency and diversity in *GIDIA*. The experiments were carried out in the Laboratory of Computer Network and Information Security at Sichuan University. We developed a grid simulation toolkit based on GridSim [10] to satisfy our needs. The simulation environment simulated users with different behaviors, resources and agents. A total of 40 computers in a grid were under surveillance. The length of selfs was 1 K. The affinity threshold was 0.7. Suppose all the mature MoAs, including the degraded memory MoAs had tolerated to those newly added self elements, and the degraded memory MoAs in evolvement into memory MoAs were prior to other mature MoAs. Every MoA carried an antibody. The mature MoAs maximum Γ was set to 256 and lifecycle λ was set to 20. 16 mature MoAs were evolved at a time. The initial number of memory MoAs was 192. The memory MoAs maximum κ was set to 256. P was set to 80. To simplify environments, we define a function $f_{generation}(t)$ of the generation of new immature MoAs. If new immature MoAs are generated, the value of $f_{generation}(t)$ is 1, 2, 3, respectively, when the survival time $T - \lambda$ is 4, 2, and 1. Otherwise, the value of $f_{generation}(t)$ remains 0.

The effects of T on the number of mature MoAs are shown in Fig.2. Given the survival time $T-\lambda=4$, 80 memory MoAs degraded into mature ones when the number of memory MoAs reached the given maximum 256 at the fourth, ninth, fourteenth, nineteenth, twenty-fourth time point, respectively. Given the survival time $T-\lambda=2$, only 32 degraded memory MoAs re-evolved into memory MoAs and other degraded memory MoAs died due to exceeding the survival time at the eighth, fourteenth, twentieth time point, respectively. 80 memory MoAs degraded into mature ones at the fourth, tenth, sixteenth, twenty-second time point, respectively. Given the survival time $T-\lambda=1$, only 16 degraded memory MoAs re-evolved into memory MoAs and other degraded memory MoAs died due to exceeding the survival time at the seventh, thirteenth, nineteenth time point, respectively. 80 memory MoAs degraded into mature ones at the fourth, tenth, sixteenth, twenty-second time point.

The generation of new immature MoAs is shown in Fig.3. Before the sixth time point, there had been new immature MoA to be generated. Given the survival time $T-\lambda=4$, new immature MoA were generated again only at the tenth, fifteenth, twentieth, twenty-fourth time point, respectively. Given the survival time $T-\lambda=2$, new immature MoA were generated again at the ninth, tenth, eleventh, fifteenth, sixteenth, seventeenth, twenty-first, twenty-second, twenty-third time point, respectively. Given the survival time $T-\lambda=1$, new immature MoA were generated again at the eighth, ninth, tenth, eleventh, fourteenth, fifteenth, sixteenth, seventeenth, twentieth, twenty-first, twenty-second, twenty-third time point, respectively.

Obviously, when the survival time of the degraded memory MoAs was short, the degraded memory MoA died soon because it could not match enough antigens in lifecycle λ. Because the degraded memory MoAs have better detection capability than mature MoAs, the death of too much degraded memory MoAs will decrease detection efficiency. Contrarily, the longer the survival time, the less new immature MoAs were generation. Of course, the number of new generated mature MoAs was small. Therefore, too large value of T will affect the diversity of *GIDIA*.

Therefore, the experiments results show that the appropriate value of T assures high detection efficiency and diversity in *GIDIA*.

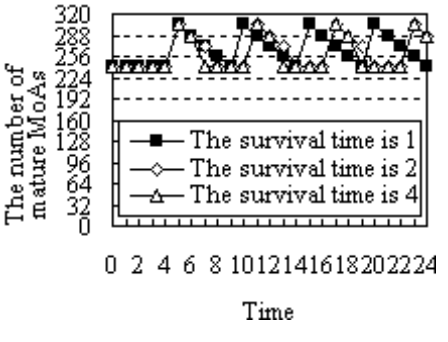


Fig. 2. Effect of T on the number of mature MoAs

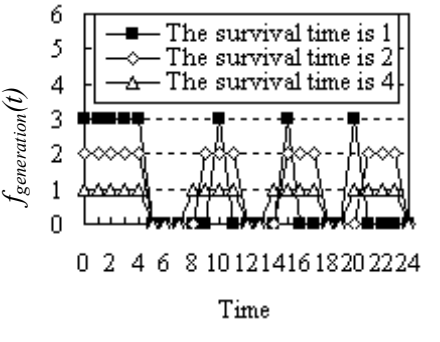


Fig. 3. Effect of T on the generation of new immature MoAs

4 Conclusions

The proposed grid intrusion detection model based on immune agent (*GIDIA*) is an active grid security technique.

In *GIDIA*, the concept of *self*, *nonself*, *antigen*, *antibody*, *agent* and *match algorithm* have been abstracted and extended. Besides, *GIDIA* introduces a new idea of dynamic memory MoAs survival, which enhances detection efficiency and assures steady detection performance. Furthermore, the mathematical models of *mature MoA* and *dynamic memory MoA survival* are presented.

The important parameter T is discussed in the theoretical analysis and the experiment results, which show that the proposed method is an effective solution to intrusion detection for grid security.

Acknowledgment

This work is partially supported by the National Natural Science Foundation of China under Grant No. 60373110, 60573130 and 60502011, the National Research Foundation for the Doctoral Program of Higher Education of China under Grant No.20030610003, the New Century Excellent Expert Program of Ministry of Education of China under Grant No. NCET-04-0870, and the Innovation Foundation of Sichuan University under Grant No.2004CF10.

References

1. Foster, Carl Kesselman, Gene Tsudik and Steven Tuecke: A Security Architecture for Computational Grids. In Proc. of Computer and Communication Security (1998)
2. S. Forrest, S. Hofmeyr, and A. Somayaji: Computer Immunology. In Proc. of Communications of the ACM, Vol. 40 (1997) 88-96
3. A. Somayaji, S. A. Hofmeyr, and S. Forrest: Principles of a Computer Immune System. In Proc. of the New Security Paradigms'97 (1997) 75-82
4. Tao Li: Compute immunology. Publishing House of Electronics Industry, Beijing (2004)
5. T. Li: An immune based dynamic intrusion detection model. Chinese Science Bulletin (2005) 2650 – 2657
6. T. Li: An immunity based network security risk estimation. Science in China Ser. F Information Sciences (2005) 557 – 578
7. T. Li: A New Model for Dynamic Intrusion Detection. Lecture Notes in Computer Science. Springer-Verlag, Berlin Heidelberg New York (2005) 72 - 84
8. D. Dasgupta: Immunity-based intrusion detection system: A general framework. In Proc. of 22nd National Information Systems Security Conference (1999) 147-160
9. Paul K. Harmer, Paul D. Williams, Gregg H. Gunsch, and Gary B. Lamont: An Artificial Immune System Architecture for Computer Security Applications. In Proc. of IEEE Transactions on Evolutionary Computation, Vol. 6 (2002)
10. M. Murshed, R. Buyya, D. Abramson: GridSim: A Grid Simulation Toolkit for Resource Management and Scheduling in Large-Scale Grid Computing Environments. In Proc. of the 17th IEEE International Symposium on Parallel and Distributed (2002)

A Novel Artificial Immune Network Model and Analysis on Its Dynamic Behavior and Stabilities*

Liya Wang, Lei Wang, and Yinling Nie

School of Computer Science and Engineering, Xi'an University of Technology
710048 Xi'an, China
wangliya66@163.com, leiwang@xaut.edu.cn

Abstract. A novel model of artificial immune network is presented at first, and then a simulative research work is made on its dynamic behaviors. Simulation results show that the limit cycle and chaos may exist simultaneously when four units are in connection, and the network's characteristic has a close relationship with the intensity of suppressor T-cell's function, B-cell's characteristics and transconductance. Besides this, with Liapunov's method, the sufficient conditions for network's stability is studied, especially for the case of system's characteristics under the condition that the helper T-cells appear as a nonlinear function.

1 Introduction

As we know, computation theories inspired from natural phenomena are being paid more and more attention in the domestic research field. Different from traditional methods, and with the condition of admitting the existent means the logical, naturally heuristic methods always try to find the order of nature discovered by now, and then regard them as the rule or principles of constructing novel computer structures or algorithms with high performance. In which, a typical example is about the study and the development in the research of artificial immune system (namely, AIS for short). Especially, computer science researchers are always interested in intelligent behaviors displayed by biology, and bring out inspirations from their structure, function and mechanisms [1]. In 1974, American Nobel prize owner N. K. Jerne advanced the famous immune network theory, who suggested that a biological immune system is in fact composed of many different kinds of immune cells [2][3]. Once the emergence of Dr. Jerne's theory, it aroused a great interest. Some years later, J. D. Farmer et al successfully used this theory for machine-learning research, which sets up an important foundation for the development of modern computation and immune mechanisms based information-processing systems [4].

In recent years, many different kinds of researches and applications of AIS have already been found such as optimization, intelligent control, pattern cognition, network security, hardware design etc. Among which, Prof. S. Forrest at University of New Mexico proposed a negation selection algorithm, which is normally used as a detector

* This research is supported by Shaanxi Province Education Department's Science Research Project under grant n°06JK224.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 83 – 91, 2006.

for the action of artificial immune acceptance, and further educed the concept of computer immune system[5][6]. Referring to existing discoveries of biologic immune system, this model is also a multi-agent system (MAS), distributive and without central control. Another Prof. D. Dasgupta at Memphis University made a deep research on immune MAS [7], and pointed out most of application problems dealt with MAS can also be done with AIS, because it is similar to MAS not only in the methods of realization but the mechanisms of operation as well [8]. Dr. Leandro Nunes de Castro at Brazil Phnom Penh National University made a comprehensive analysis on both the artificial immune clonal selection and the artificial immune network (AIN, which is called aiNet by authors) [9][10]. aiNet is an incompletely connected graph, and anyone of its nodes can be regarded as an antibody. Dr. J. Timmis at Kent University in Britain proposed a machine-learning model based on AIS with finite resource [11][12], in which, AIS is used as a tool to abstract knowledge from experience and then for new problem solving. Dr. J. A. White at IBM proposed a novel pattern classifier and recognition method in accordance with the immune mechanism that an antibody identifies and classifies an antigen during the process of contact [13].

For natural immune system, it is a parallel, distributed and self-adaptive information-process system. From microcosmic view on this system's components, it is made up of many different kinds of cells, organs and organizations which interact and collaborate with each other so as to complete a target task. Usually, B-cells take an important role inside; however, their function would not be implemented without cooperation from other materials or various T-cells, such as T_H and T_S. On the other hand, there are various and interplayed affinities among B-cells. Additionally, such kind of relations may take some complex and dynamic changes due to cell's death or clonal procreation. Since 1997, some AIN models have been proposed for simulating the characters of the natural [14][15]. Although these models are distributed, self-adaptive, and suitable for some purpose of complex applications, over a period of practice, they are found not satisfied at diversity maintenance, or distributed processing, or fault tolerance of information, or dynamic learning, or self-recovery. There are perhaps two reasons inside, firstly, it is a gradually deepening process of our recognition and understanding to nature; and secondly, our research in this area is insufficient, which certainly brings about some difficulties to our current work.

From analysis above, it can be seen that the immune mechanisms is very import to construct a model whether on its organization or structure. In addition, the stability of a system or a network is the basis of further theoretical analysis and applications. However, this is usually ignored in some degree. Due to lack of powerful theoretical work, current research on AIS is mostly limited to some simulations of simple functions of biology immune system. Deep-going research has not been done yet on how to construct a network which is provided with most important features of natural systems. In other words, such kind of examples have been seldom seen that integrate equilibrium point, limit cycle and chaos into one system or network. Therefore, novel AIN models are still needed whose behaviors should: 1, integrate equilibrium point, limit cycle, chaos and other characteristics together, which could occur respectively with parameter's variation in the same model; 2, have a simple structure and convenient realization; and 3, be provided with a wider versatility.

2 The Model of Artificial Immune Network

As shown above, there are already some AIN models such as aiNet, dynamic immune network, multiple-valued AIN, et al. Among which, the model based on immune response by Zhang Tang imitates some key features of natural immune system, such as the reaction between B-cells and T-cells. In detail, a population composed by suppressor T-cells (T_S, in the suppressing layer of the network) is strict with monotonously increasing or decreasing nonlinear relationship for its saturation. However, besides saturate, interactions between excitement and suppression also frequently occur in body's physiological process, which means AIN's nonlinear unit can be not only monotonous but also non-monotonous as well. Based on this consideration, it is necessary to survey if there are some features missing for constructing our artificial models. In fact, there is another population composed of assistant T-cells (T_H) in the natural immune system, which simulates the function of system's memory, and meanwhile, T_H cells can cooperate with T_S, so as to make the system appear complex and non-monotonous.

The function of cooperation between T_S and T_H sets up a feedback channel transferring information from the output layer (antibodies) to the input (antigens). If there have already been some antibodies corresponding to a special kind of antigens, other similar antigens will not activate the immune network any further when invading the system. In this case, the system appears suppressant. On the contrary, if there are some new antigens invading the system, due to no corresponding antibodies existing in the output layer, they will activate the immune network immediately and obviously, and now, the system appears excited.

In view of the above process, it is necessary to perfect our current AIN models, so as to reflect the cooperation mechanisms among B-cells and different kinds of T-cells. For the new model design, because the population of B-cells is usually in a stable state, so we can also regard it as the core of network structure, and mainly performing the task of information-processing. Under the assistant function of T_H cells, the population of antibodies generated by B-cells takes a dynamic role of either enhancing B-cell's ability of information-processing, or making some information-preprocessing before B-cell's action. On the other hand, there are two different kinds of channels in the network, that are respectively the information-processing channel, namely a forward channel composed by B-cells and antibodies; and the information-transforming channel, namely a backward channel by T-cells. In which, the function of cooperation between T_S and T_H sets up a feedback channel from the output layer (antibodies) to the input (antigens) that can transmit a status information of the system. In this case, if there already exist some antibodies to a special kind of antigens, this system will not generate the same new antibodies any more. However, it is necessary to point out that the function and the power of antibodies generated by self or non-self B-cells are different during the system's information feedback. In other words, even if the system does not produce antibodies for a special purpose, it is possible that the system still generates some in order to maintain its normal status. This possibility maybe provides some diversity due to itself or non-itself reasons. Based on these

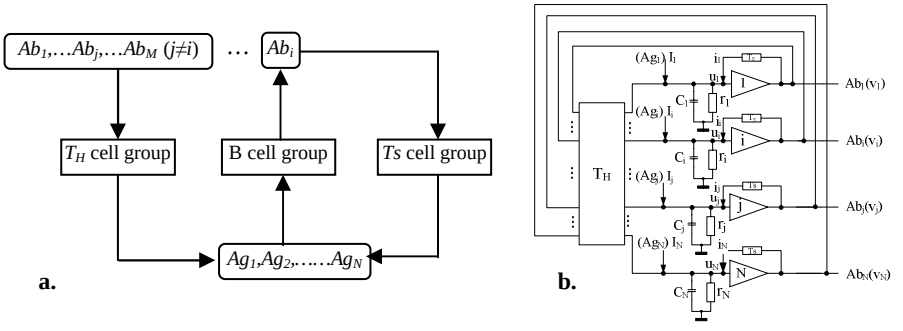


Fig. 1. The schematic structure of an artificial immune network model, in which, a, the immune response network in view of nature, and b, the engineering system model in accordance with the immune response network on its left

considerations above, an AIN model and its corresponding structure are given as shown in Fig.1.

In the model above, transconductance matrix, T_H takes a role as collecting information from other immune units. In which, T_{ij} is defined as action or connection intensity to antigen Ag_j, and T_S is shown as a nonlinear voltage-controlled units denoting the direct action of antibody Ab_i to its corresponding antigen Ag_j. Therefore, its differential equation is available from system topology configuration corresponding to Fig.1 which can be expressed as followed:

$$\begin{cases} c_i \dfrac{du_i}{dt} = \displaystyle\sum_{j=1, j\neq i}^{M} T_{ij}\left(Ab_j - u_i\right) - \frac{u_i}{r_i} + i_{ii} + Ag_i \quad i = 1, 2, \cdots, N \\ Ab_i = B_i(u_i) \end{cases} \tag{1}$$

where T_{Si} suggests a nonlinear voltage-controlled unit denoting the feedback characteristic of antibody generated by a single white blood cell to its antigen, that is $i_{ii} = G_i(u_{ii}) = G_i(Ab_i - u_i)$. Besides this and for the convenience, we let $R_i = r_i + 1 \Big/ \sum T_{ij}$, then the equation above can be changed as:

$$\frac{du_i}{dt} = \sum_{j=1, j\neq i}^{M} \frac{T_{ij}}{c_i} B_j(u_j) - \frac{u_i}{c_i R_i} + G_i(B_i(u_i) - u_i) + Ag_i \quad i = 1, 2, \cdots, N \tag{2}$$

In the equation above, $B_i(\cdot)$ is defined as a continuous function, and $G_i(\cdot)$ is a continuous differentiable monotonous function. Obviously, equation (2) will be common form of Hopfield network when $i_{ii} \equiv 0$ and $T_{ij}=T_{ji}$, as well as $B_i(\cdot)$ belongs to continuous differentiable monotonous function. Besides this, the network will be a continuous asymmetrical form when $i_{ii} \equiv 0$ and $T_{ij} \neq T_{ji}$. Therefore, equation (2) suggests a more general situation.

Here we limit the population of T_S cells in the system to a saturated situation, in other words, when a kind of antibodies generated by B-cells reach achieve to a certain

level, their function to these antigens will not strengthen any more. Therefore, for the continuous function $B_i(\cdot)$ simulating the characteristics of B-cells, it shows the nonlinear part of this unit which has the following five types: 1, monotone-saturation type, as a majority units of this network model, and from the dynamics point of view, if there exists an isolated equilibrium point, then it must be exclusive; 2, excitement type, whose output of leucocytes is a monotonously increasing function of the interior status; 3, suppression type, contrary to excitement type, whose output of leucocytes is a monotonously decreasing function; 4, excitement-suppression type, whose output of leucocytes firstly rises with the increase of its interior status, and then began to descend when reaching a certain degree; and 5, suppression-excitement type, contrary to the excitement-suppression type, it firstly descends with the increase of its interior status and then rises when arriving at a limit. To be clearer, all of these models are illustrated in Fig.2. In this part of the follow-up we can see that if these types construct a large-scale network, its dynamics characteristics might be a more complex character and different stability.

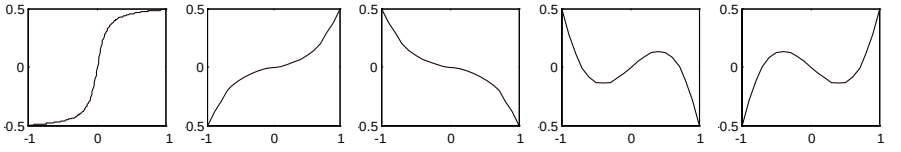


Fig. 2. Five types that a B-cell could be, a, monotone-saturation type; b, excitement type; c, suppression type; d, excitement-suppression type; and e, suppression-excitement type

3 Simulation on AIN's Dynamic Behaviors

With digital simulation on the above models, it can be seen that this model is provided with some very complex characteristics when the network is under three- or four-unit's connection. In this model, the nonlinear characteristics of $B_i(\cdot)$ are given by:

$$B_i(u_i) = a_i u_i + b_i u_i^3 \tag{3}$$

When $a_i > 0$, $b_i < 0$, it means that this model belongs to the excitement type, and oppositely, $a_i < 0$, $b_i < 0$, means the suppression type, and $a_i > 0$, $b_i < 0$, the excitement-suppression type. On the other hand, the characteristics of T_S are defined as:

$$i_{ii} = G_i(u_{ii}) = c_i k_i \frac{1 - e^{-u_{ii}}}{1 + e^{-u_{ii}}} \tag{4}$$

where k_i suggests a saturated feed-back value, c_i means the feed-back intensity, and especially, $c_i=0$ indicates no-feedback and $c_i=1$ is whole-feedback. In order to reflect the more comprehensive characteristics the network behaves at different conditions,

we observe it under four-unit's connection to examine its characteristics (the chaotic phenomenon is not found when the network is under three-unit's connection).

3.1 Influence of the Transconductance Matrix T_H on System's Dynamics

When the Transconductance matrix T_H is asymmetrical, for example in the case of four-units connection network, when $T_H = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 0 & 1 & -0.1 \\ -0.1 & 1 & 0 & 1 \\ 1 & -0.1 & 2 & 0 \end{bmatrix}$, and a_1=5, b_1=-1; a_2=8, b_2=-2; a_3=10, b_3=4; a_4=6, b_4=-2. With different initial conditions, there usually exist some stable limit-cycles, as shown in Fig.3(a). However, when the transconductance matrix T_H is symmetrical, for example when $T_H = \begin{bmatrix} 0 & 1 & -0.1 & 1 \\ 1 & 0 & 1 & -0.1 \\ -0.1 & 1 & 0 & 1 \\ 1 & -0.1 & 1 & 0 \end{bmatrix}$, and the rest are as same as the above, some asymptotically stable limit-cycles appear as shown in Fig.3(b).

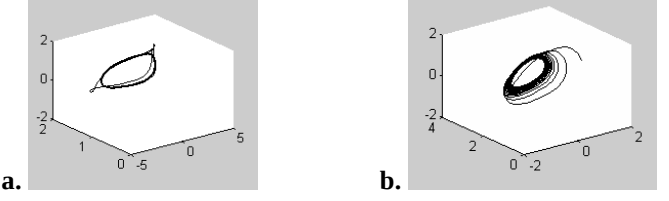


Fig. 3. Influence of the Transconductance matrix T_H on dynamic locus, a, Transconductance matrix is asymmetrical, and b, Transconductance matrix is symmetrical

3.2 Influence of B-Cell's Characteristics on System's Dynamics

When k_i=1, c_i=1, namely the whole-feedback scenario, and a_1=8, b_1=-1; a_2=6, b_2=-2; a_3=10, b_3=-9; a_4=6, b_4=-4, the nonlinear characteristics of unit belong to excitement-suppression type, the results of simulation show that equilibrium point only may occur, which are shown in Fig.4(a). When a_1=8, b_1=-1; a_2=13, b_2=-2; a_3=6, b_3=-3; a_4=10, b_4=-4, namely the nonlinear characteristics of cell belong to excitement-suppression type all the same, then the corresponding limit cycle occurs as shown in Fig.4(b). By only changing a_4 from 6 to 1; b_4 from -4 to 2 and thereby turning B-cell's characteristics into the excitement type, chaotic phenomenon now appears as shown in Fig.4(c).

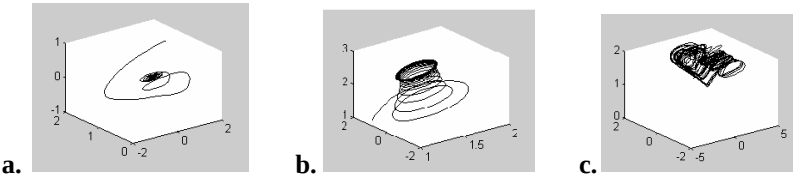


Fig. 4. Influence of B-cell's characteristics on system's dynamics, a, the equilibrium point of a four-order network; b, the limit cycle; and c, the chaotic phenomenon

3.3 Influence of the Suppressor T-Cell's Characteristics on System's Dynamics

For convenient comparison, the experiment is based on the parameters of when chaotic phenomenon appears, only changing the feed-back intensity *c*. Dynamic locus can be obtained as shown in Fig.5 with c_i=0, c_i=0.5, c_i=1.0, c_i=1.2 respectively. It can be seen from Fig.5 that if c_i =0 (i.e. no-feedback), only a equilibrium point occurs; if c_i =0.5, the dynamic locus changes to a limit circle; if c_i =1, chaotic phenomenon appears in dynamic locus; and if c_i =1.2, the track tends to a limit circle after running chaotically for a while. Finally it should be pointed out that when nonlinear character is of monotonous-increasing type, there is no presence of limit cycle, irrespective of the choice of whatever parameters. In addition, it seems that there is even no sign of spiral approach of dynamic locus towards equilibrium point. These outlined above show: the networks structured by monotonous increasing satiation units have a simple characteristic, which is unable to analogize the memory and response process of body's immune system.

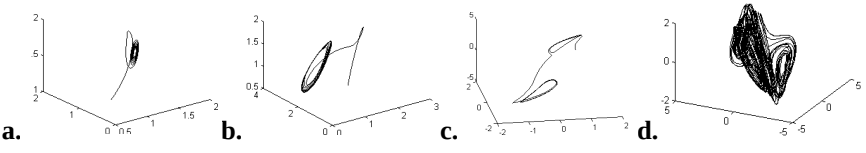


Fig. 5. Influence of the feed-back intensity on dynamic locus, in which, a, c_i=0; b, c_i=0.5; c, c_i=1; and d, c_i=1.2

4 Analysis on the Network's Stability

Based on the analysis above, it can be seen that the network is able to eliminate antigen's action through activating antibody's multiplication, and finally tends to a stable status. In other words, it is a gradual process that the system reaches its equilibrium, which provides the conditions for utilizing Lyapunov's methods to analysis this model's stability. In detail, for the system described as equation (2), we presume the following two properties exist in the active function $B(\cdot)$ of B-cells and the reactive function $G(\cdot)$ of the antibody with the assistance of helper T_S-cells. Here we suppose $u_{ii} = Ab_i - u_i$.

Theorem 1: According to the property of reaction from antibodies (through the helper T_S-cells), $G_i(\cdot)$ in the real number set R is a monotonously increased function, and there exist a series of constant number α_i, β_i, where $0<\alpha_i<\beta_i<\infty$ (i = 1, 2,…,n), which makes the following equation satisfied for any u, $v \in R$, and $u \neq v$.

$$\alpha \leq \frac{\mathrm{G}(u)-\mathrm{G}(v)}{u-v} \leq \beta \tag{5}$$

Due to the limitation of paper length, the detailed proof process is omitted here, and we let $I = diag(\alpha_1, \alpha_2, ..., \alpha_n)$ and $L = diag(\beta_1, \beta_2, ..., \beta_n)$.

Theorem 2: For the active function $B_i(u_i) = a_i u_i + b_i u_i^3$, there exists a constant m_i, $0<m_i<\infty$ (i=1,2,…,n), which makes the following equation satisfied for any u, $v \in R$, and here we let $M = diag(m_1, m_2, ..., m_n)$.

$$0 < \frac{B_i(u)-B_i(v)}{u-v} < m_i \tag{6}$$

For the discussion of system's stability, it is also necessary to introduce two definitions from the reference [16], which are respectively:

Definition 1: If all principal forms of a matrix A are positive, then we call the matrix A belongs to a class of matrixes P, and mark it as $A \in P$.

Definition 2: For a matrix A, if there exists a diagonal matrix $\alpha = \mathrm{diag}(\alpha_1, \alpha_2, ..., \alpha_n) > 0$ such that $[\alpha A]^s > 0$, in which $[\alpha \mathrm{A}]^s = (A^T \alpha + \alpha \mathrm{T})/2$, then we call A is Liapunov's diagonal stable, and mark it as $A \in LDS$.

From the above theorems and definitions, we can conclude the following theorem.

Theorem 3: If the functions $G(\cdot)$ and $B(\cdot)$ satisfy the theorems 1 and 2 respectively, and $-A + L'M^{-1} \in LDS$, then for any $I \in R^n$, there is not only an exclusive equilibrium point in the system, but this system is global stable with exponential level as well.

5 Conclusion

Based on the statements above, we firstly conclude that AIN's dynamics has a close relationship with its transconductance, features of B-cells and the function's intension of suppressor T-cells. From the analysis on simulation results, it can also be seen that system's dynamics with symmetrical transconductance is more complex the asymmetrical. Especially, when some B-cells belong to the excitement-suppression type and the rest are the excitement type, chaos may appear.

Secondly, for the system described by equation (2), with theorem 1, we can find out that, 1, if B-cells belong to the excitement type, the network is globally stable with exponential level; 2, if B-cells are the suppression type, it is not enough to forecast whether the system is globally stable or not; and 3, if B-cells are the excitement-suppression type, then the system is stable under the condition of existing u_1 and u_2,

which satisfies $-\sqrt{-a_i/3b_i} \le u_1 < u_2 \le \sqrt{-a_i/3b_i}$,or if B-cells are the suppression-excitement type, then it is stable under the condition of $\sqrt{-a_i/3b_i} \le u_1 < u_2$ or $u_1 < u_2 \le -\sqrt{-a_i/3b_i}$.

References

1. Klarreich, E.: Inspired by Immunity. Nature 415(31) (2002) 468-470
2. Jerne, N. K.: Towards a Network Theory of the Immune System. Annual Immunology 125C (1974) 373-389
3. Jerne, N. K.: The Generative Grammar of the Immune System. In: Wigzell, H. (eds.): Presentation Speech to Nobel Prize in Physical or Medicine (1984) http://nobelprize.org/medicine/laureates/1984/presentation-speech.html
4. Farmer, J. D., Packard, N. H., Perelson, A. S.: The Immune System. Adaptation, and Machine Learning. Physica 22D (1986) 187-204
5. Chao, D. L., Davenport, M. P., Forrest, S., Perelson, A. S.: A stochastic model of cytotoxic T cell responses. Journal of Theoretical Biology 228(2) (2004) 227-240
6. Esponda, F., Forrest, S., Helman, P.: A formal framework for positive and negative detection. IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics 34(1) (2004) 357-373
7. Dasgupta, D., Gonzalez, F.: An Immunity-Based Technique to Characterize Intrusions in Computer Networks. IEEE Transactions on Evolutionary Computation 6(3) (2002) 156-162
8. Gomez, J., Gonzalez, F., Dasgupta, D.: An Immuno-Fuzzy Approach to Anomaly Detection. In: Proceedings of IEEE International Conference on Fuzzy Systems (2003) 1219-1224
9. De Sousa, J. S., de Gomes, L., Bezerra, G. B., de Castro, L. N., Von Zuben, F. J.: An Immune-Evolutionary Algorithm for Multiple Rearrangements of Gene Expression Data. Genetic Programming and Evolvable Machines 5 (2004) 157-179
10. De Castro, L. N., Von Zuben, F. J.: Learning and Optimization Using the Clonal Selection Principle. IEEE Transactions on Evolutionary Computation 6(3) (2002) 239-251
11. Timmis, J., Boggess, L., Watkins, A.: Artificial Immune Recognition System (AIRS): An Immune Inspired Supervised Machine Learning Algorithm. Genetic Programming and Evolvable Machines 5(1) (2004) 51-58
12. Watkins, J. Timmis. Exploiting Parallelism Inherent in AIRS, an Artificial Immune Classifier. Lecture Notes in Computer Science, Vol.3239. Springer-Verlag, Berlin Heidelberg (2004) 427-438
13. White, J. A., Garrett, S. M.: Improved Pattern Recognition with Artificial Clonal Selection. In: Proceedings of the 2nd International Conference on Artificial Immune Systems 181-193
14. Ishiguro, K., Kondo, T., Watanabe, Y.: Emergent Construction of Artificial Immune Networks for Autonomous Mobile Robots. In: Proceedings of IEEE International Conference on System Man and Cybernetics (1997) 1222-1228
15. Tang, Z., Yamaguchi, T., Tashima, K., Ishizuka, O., Tanno, K.: Multiple-Valued Immune Network Model and Its Simulations. In: Proceedings of 27th IEEE International Symposium on Multiple-Valued Logic (1997) 233-238
16. Tan, X. H., Zhang, J. Y., Yang, Y. R.: Study on Global Exponential Stability of Neural Networks and Its Convergence Estimate. Journal of Electronics and Information Technology 25(10) (2003): 1361-1366

Immune Algorithm Optimization of Membership Functions for Mining Association Rules

Hongwei Mo[1], Xiquan Zuo[2], and Lifang Xu[1]

[1] Automation College, Harbin Engineering University,
150001 Harbin, China
{honwei2004}@126.com,{mxlfang}@sohu.com
[2] College of Computer Science and Technology, Beijing University of Posts and Telecommunications, 100080 Beijing, China
{xinguanz}@sina.com

Abstract. In the paper, immune algorithm(IA) is proposed for optimizing membership function of fuzzy variables for mining associate rules. It is used in network detection to testify its efficiency in such mining task, including maximizing the similarity between normal association rule sets while minimizing the similarity between a normal and an abnormal association rule set. Experiment results show that IA-optimization based fuzzy logic system can improve the performance of mining associate rules in network intrusion.

1 Introduction

Fuzzy logic is a useful tool for modeling complex systems and deriving useful fuzzy relations or rules [9]. GA had been used to optimize the membership functions of the fuzzy sets used by fuzzy systems[15],[16]. [12] has developed a method for mining fuzzy association rules for intrusion detection. This method is a fuzzy extension of the techniques used by Lee[10],[11], in which one compares the similarity of association rules mined from audit data when there are no intrusions with rules mined from audit data when there are intrusions.

In this paper, we report on the use of immune algorithm to tune the membership functions of the fuzzy variables used to mine the fuzzy association rules in order to improve the performance of the intrusion detection system and testify the use of immune algorithm in such mining task. The goal is to maximize the similarity of sets of normal rules and minimize the similarity of normal and abnormal rule sets. Non-intrusion is considered here as a normal situation and intrusion as an abnormal situation. Similarly, rules mined when there are no intrusions comprise a description of the normal situation while rules mined in the presence of an intrusion describe an abnormal situation. The remainder of this paper is organized as follows. Section 2 presents a brief overview of AIS, and related work about AIS in the field of data mining and computer intrusion detection. Section 3 describes immune algorithm. Section 4 reports computational results. Finally, section 5 presents the conclusions and future research directions.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 92 – 99, 2006.

2 Artificial Immune Systems and Related Work

Immune system can be divided into innate immune system and adaptive immune system. Innate immune system is composed mainly of phagocytes and the complement system(blood proteins) It is the first line of defense that works rapidly.

Adaptive immune system is composed mainly of lymphocytes, such as B-cells and T-cells. B-Cells are generated and developed in the Bone Marrow. T-Cells are generated in the Bone Marrow and developed in the Thymus. They can generates specific antigen receptors on their surface (clonal selection of B-Cells and T-Cells). Adaptive immune system is capable of fine-tuning the cell receptors(maturation of B-Cells) and confers resistance against future infections (memory of B-Cells and T-Cells).AIS consist of methods that are inspired by the biological immune system and designed to solve real-world problems [2]. This work focuses on one kind of AIS inspired by the mechanisms of the biological immune system. We now briefly review related work. Immune algorithm had been used in solving optimization such as n-TSP[4]. Another standard application is function optimization [3],[8]. AIS have been applied successfully in anomaly based computer network intrusion detection[5][7]. It has been suggested that an AIS based on the clonal selection principle, called CLONALG, can be used for classification in the context of pattern recognition[2], although originally proposed for other tasks. An AIS for discovering IF-THEN rules is proposed in [6]. A fuzzy AIS is proposed in [14]. However, that work addresses the task of clustering, which is very different from the task of optimization addressed in this paper.

3 Immune Algorithm

This section introduces the basic principle and process of immune algorithm.

3.1 Process of Immune Algorithm

Figure 1 shows the adaptive immune system. Immune algorithm inspired by such process and its steps are shown in Figure 2. First step of immune algorithm is to initialize the problem environment. Second step is to generate some candidate solutions.

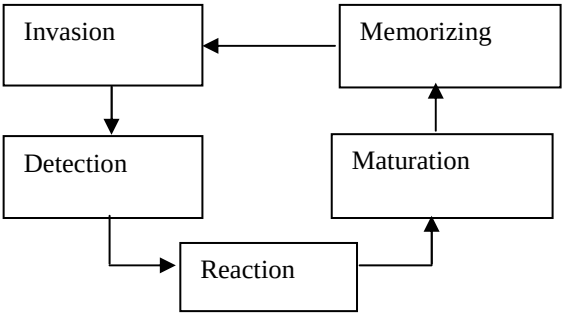


Fig. 1. Adaptive process of immune system

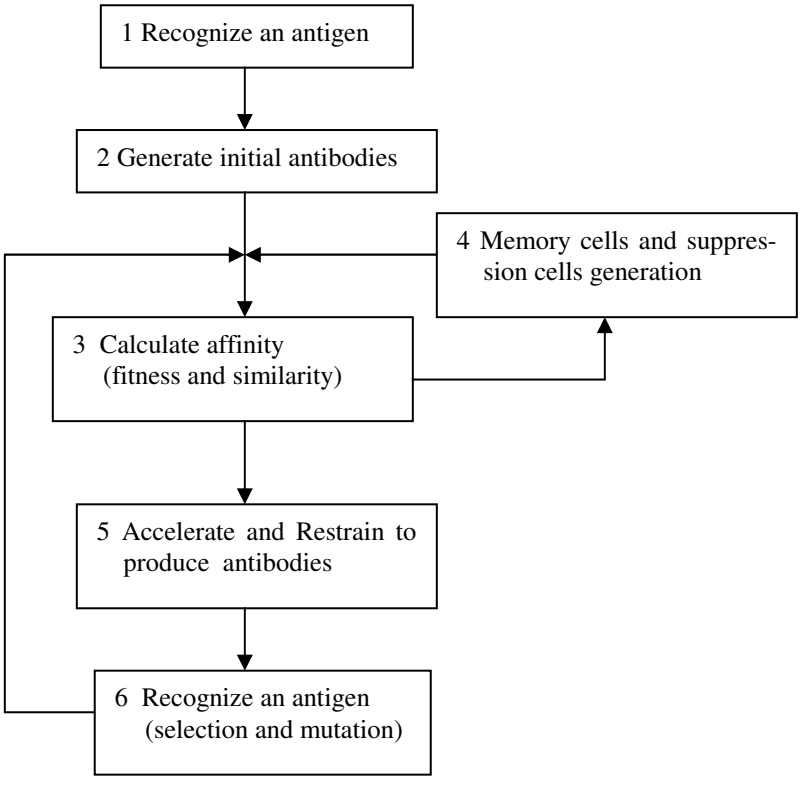


Fig. 2. The process of immune algorithm

Third is to calculate evaluation value(like GA) and similarity value of each solution. Fourth is to memory some important solutions on suppression cells and memory cells The memory on suppression cell is used to change search scope effectively. The memory on memory cell is to keep the candidate solutions. Fifth is to modify the fitness landscape based on the similarity value of each solution. Sixth is to reproduce candidate solutions using selection, crossover and mutation operators.

For the antibodies that survived in last step and for the newly produced antibodies,N/4 sets of antibodies pairs are selected by allowing duplications(N is the total number of antibodies). However, it is assumed that antibodies with higher fitness values are more likely to be selected. For the antibodies that become a pair, N/2 new antibodies are produced by crossover. For the antibodies produced, the genes are changed using a preset mutation probability and mutation operation method. However, the operation method and the probability of crossover and mutation are set arbitrarily. Until the preset final generation is reached, it returns to step 3.

3.2 Data Structure

In this initial work, each fuzzy variable is defined by 3 fuzzy sets. We use the standard Z, Π, S functions as the membership functions for these fuzzy sets. Each of

these functions has two parameters as shown in Figure 3. An antibody/antigen gene in the fuzzy logic/immune algorithm is the set of 6 parameters that are used to define the standard membership functions of a fuzzy variable(Figure 3).In turn, an antibody is a string of such genes where each gene represents a different fuzzy variable. For the example of intrusion detection task, the fuzzy variable SN,FN and RN are the number of SYN,FIN and RST Ags appearing in TCP packet headers during last 2 seconds. The fuzzy variable PN is the number of different destination ports during network traffic are thought to be diagnostic for network intrusions[13].

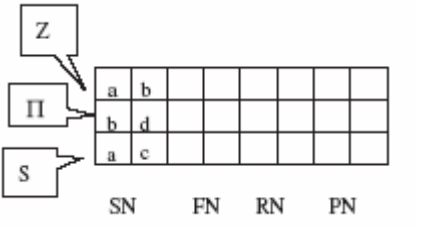


Fig. 3. The process of immune algorithm

3.3 Fitness Functions

The fitness functions for the immune algorithm is based on the similarity of rule sets mined from different sets of audit data. [12]defines the similarity between two fuzzy association rules and between tow fuzzy rules sets. An association rule is of the form: $X \to Y, c, s$,where X and Y are two disjoint item sets $(X \cap Y = \phi)$, s is the support for the association rule, and c is the confidence for the rule. The support for a rule is the percentage of transactions in which X and Y appear in the same transaction: $s = n^{'} / n$ ($n^{'}$ is the number of transactions that contain both X and Y, n is the total number of transactions). The confidence in the rule is the percentage of the transactions containing X that also contain: $Y : c = n^{'} / n^{''}$ ($n^{''}$ is the number of transactions that contains X). Algorithms for finding association rules in transaction data use two parameters called the minimum support threshold and the minimum confidence threshold [1]. Only rules with *s* and *c* values above the thresholds are mined from the data. Given two association rules $R1 : X \to Y, c, s$ and $R2 : X^{'} \to Y^{'}, c^{'}, s^{'}$,if $X = X^{'}$ and $Y = Y^{'}$ then $R1$ and $R2$ is

$$similarity(R1, R2) = \max(0, 1 - \max(\frac{|c - c^{'}|}{c}, \frac{|s - s^{'}|}{s})) . \quad (1)$$

otherwise, $similarity(R1, R2) = 0$. The similarity between two rule sets $S1$ and $S2$ is

$$similarity(S1,S2) = \frac{s}{|S1|} * \frac{s}{|S2|}. \tag{2}$$

where $s = \sum similarity(R1,R2), \forall R1 \in S1, \forall R2 \in S2$, $|S1|$ and $|S2|$ are the total number of rules in $S1$ and $|R2|$,respectively, where $|S1|$ and $|S2|$ are the total number of rules in $S1$ and $S2$,respectively. Three different sets of audit data were available for testing the fuzzy logic genetic algorithm: a normal data set with no intrusions and two "abnormal" data sets with different types of intrusions. The normal data set was partitioned into two sets. One partition of the normal audit data is called the reference data and one set is called normal data. The following five fitness functions have been designed and tested:

$$F_1 = \frac{S_{rn}}{S_{ra1}} * \frac{S_{rn}}{S_{ra2}}. \tag{3}$$

$$F_2 = \frac{S_{rn}}{S_{ra1}} + \frac{S_{rn}}{S_{ra2}}. \tag{4}$$

$$F_3 = 2S_{rn} - S_{ra1} - S_{ra2}. \tag{5}$$

$$F_4 = \frac{S_{rn}}{S_{ra1}}. \tag{6}$$

$$F_5 = S_{rn} - S_{r1}. \tag{7}$$

where S_{rn} is the similarity between a reference rule set and the ‘normal’ rule set, S_{ra1} and S_{ra2} are respectively the similarity between the reference rule set and abnormal rule set 1 and between the reference rule set and abnormal rule set 2. In practice, a small constant has been added to all denominators in the above fitness functions to avoid division by zero. Functions (3)-(5) will drive a co-evolution of the system's sensitivity to the two types of intrusions, because the similarities between the reference rule set and the two anomaly rule sets must be minimized at the same time. Functions (6)-(7) are designed to test the versatility of our intrusion detection approach, i.e., a system trained with only one type of intrusion is used for detecting another type of intrusion.

4 Experimental Results

The network intrusion detection system has been trained and tested using three sets of network traffic data downloaded from http://iris.cs.uml.edu:8080. These three data sets were collected by tcpdump. The first data set, baseline, was collected in normal

situations (no network intrusion), the second data set, abnormal1, was collected when "IP spoofing" intrusions were simulated, and the third data set, abnormal2, was collected when "port scanning" intrusions were simulated. Four diagnostic Ags (SN, FN, RN, PN) were extracted from the raw data sets. The data set Baseline was divided into two partitions, one of which will be called reference and the other normal. Reference was used to mine the reference association rule set that represents the normal situations, while normal, abnormal1 and abnormal2 were used to mine new association rule sets to be compared with the reference rule set. The goal is to use the IA to derive membership functions that maximize normal similarity, and minimize both abnormal1 similarity, and abnormal2 similarity. Normal similarity, abnormal1 similarity, and abnormal2 similarity are short names for the similarity between the reference association rule set mined from reference and one of the three rule sets mined from normal, abnormal1, and abnormal2.

Table 1. Similarities of rule sets to the reference rule set before optimization and using different fitness functions

State	Original	F1	F2	F3	F4	F5
Normal	0.74	0.03	0.08	0.06	0.04	0.04
Abnormal	0.31	0.04	0.06	0.06	0.05	0.03
Abnorma2	0.32	0.00	0.01	0.02	0.05	0.03

Table 1 shows the similarities before and after optimization with all five fitness functions. In all cases the similarities of the references rule set and normal rule set are increased while the similarities of the reference rule set and abnormal sets are decreased. The results using fitness functions F4 and F5 indicate the generality of our detection approach, since the system was trained in those experiments with only one type of intrusion, but it was able to recognize another type of intrusion. As a matter of fact, abnormal1 similarity and abnormal2 similarity were found close to each other in all experiments, regardless of the fitness function used or the parameter settings. In addition to the effect of the fitness functions, the settings of minimum support and minimum confidence obviously influence the system performance.

Table 2. Similarities between the reference rule set and the 3 test rule sets (normal, anormal1, abnormal2) before and after IA optimization using fitness function F4

State	0.6	0.8
Normal	0.761	0.82
Abnormal	0.225	0.04
Abnorma2	0.224	0.04

Table 2 presents the results of an experiment comparing the similarity of the three rule sets and the reference set before and after the IA was used to optimize the membership functions. Fitness function F4 was used in this experiment. In the pair of bars, the former one is before IA optimization and the later is after IA optimization.

Table 3. Different minimum supports lead to different sets of anormal1 similarities

Min.Support	F1	F2	F3	F4	F5
0.6	0.141	0.24	0.29	0.32	0.34
0.8	0.045	0.044	0.04	0.044	0.04

Table 3 shows the effect of two different minimum support thresholds when abnormal set 1 is compared to the reference set for all fitness functions. The lower minimum support value results in a larger number of association rules with a greater degree of variation. In the table, one is minimum support=0.6 and another is minimum support=0.8.

5 Conclusions

A IA-optimized fuzzy logic system has been experimentally shown to have an improved performance for mining associate rules in network intrusion, i.e., maximizing similarity between a normal and an abnormal association rule set. It shows that IA can be used in such task. Further experiments will be conducted to determine the effects of additional variations in the IA parameters and algorithm. It can also be used to optimize the membership functions for fuzzy serial frequency episodes, and to tune other association rule mining parameters such as minimum support and minimum confidence. The further application of IA in data mining will be researched, too.

Acknowledgement

The work is supported by National Nature Science Foundation of China, No.6030 5007 and Foundation for Persons Studying abroad Homecoming of Heilongjiang Province,No.LC05C03.

References

1. Agrawal, R., Srikant, R.: Fast Algorithms for Mining Association Rules. Proceedings of the 20th International Conference on very Large Databases. San Francisco. CA:Morgan Kaufmann (1994)
2. Castro, L. N., Timmis, J.: Artificial Immune Systems: A New Computational Intelligence Approach. Springer-Verlag, Berlin Heidelberg (2002)
3. Cutello, V., Nicosia, G.:An Immunological Approach to Combinatorial Optimization Problems. In: F.J. Garijo, J.C. Riquelme, M. Toro (Eds.): Advances in Artificial Intelligence IBERAMIA. Lecture Notes on Artificial Intelligence, Vol. 2527. Springer-Verlag, Berlin Heidelberg (2002) 361-370
4. Endoh, S.,Toma, N.,Yamada, K.:Immune Algorithm for n-TSP. Proceedings of the IEEE Systems,Man and Cybernetics Conference (1998) 3844-3849
5. Hofmeyr, S., Forrest, S.:Architecture for an Artificial Immune System. Evolutionary Computation, Vol. 8(4). (2000) 443-473

6. Gonzales, F.A., Dasgupta, D.: An Immunogenetic Technique to Detect Anomalies in Network Traffic. Proceedings of Genetic and Evolutionary Computation. San Mateo,Morgan Kaufmann (2002) 1081-1088
7. Kephart J. A Biologically Inspired Immune System for Computers. Proceedings of Artificial Life.(1994)130-139
8. Kelsey, J., Timmis, J.:Immune Inspired Somatic Contiguous Hypermutation for Function Optimisation. Proceedings of Genetic and Evolutionary Computation Conference (GECCO). Lecture Notes in Computer Science, Vol. 2723. Springer-Verlag, Berlin Heidelberg New York (2003) 207-218
9. Kuok, C., A. Fu, M. Wong. Mining Fuzzy Association Rules in Databases. SIGMOD Record Vol. 17(1). (1998) 41-6
10. Lee, W., Stolfo, S.: Data Mining Approaches for Intrusion Detection. Proceedings of the 7th USENIX Security Symposium. Berkeley, Usenix Association (1998)
11. Lee, W., Stolfo, S.,Mok.: A Data Mining Framework for Building Intrusion Detection Models.Proceedings of the 1999 IEEE Symposium on Security and Privacy.Oakland, CA(1999)
12. Porras. P., Valdes A. :Live Traffic Analysis of TCP/IP Gateways. Proceedings of the ISOC Symposium on Network and Distributed Systems Security (1998)
13. Nasaroui, O., Gonzales, F., Dasgupta, D.: The Fuzzy Artificial Immune System:Motivations, Basic Concepts, and Application to Clustering and Web Profiling. Proceedings of IEEE International Conference on Fuzzy Systems (2002) 711-716
14. Tang, K. S., Man, K. F., Liu, Z. F., Kwong, S.: Minimal Fuzzy Memberships and Rules Using Hierarchical Genetic Algorithms. IEEE Transactions on Industrial Electronics, Vol. 45(1) (1998) 162-69
15. Wang, W. D., Susan, M. B.: Genetic Algorithm Optimization of Membership Functions for Mining Fuzzy Association Rules. International Joint Conference on Information Systems, Fuzzy Theory and Technology Conference (2000)

Immune Clonal MO Algorithm for ZDT Problems

Ronghua Shang and Wenping Ma

Institute of Intelligent Information Processing, P.O. Box 224, Xidian University, Xi'an, 710071, P.R. China
shangronghua1980@163.com

Abstract. In this paper, we introduce a new multiobjective optimization (MO) algorithm to solve ZDT test problems using the immune clonal principle. This algorithm is termed Immune Clonal MO Algorithm (ICMOA). In ICMOA, the antibody population is split into nondominated antibodies and dominated antibodies. Meanwhile, the nondominated antibodies are allowed to survive and to clone and the *nonuniform mutation* is adopted. Two metrics proposed by K. Deb et al. are adopted to measure the extent of convergence to a known set of Pareto-optimal solutions and the extent of spread achieved among the obtained solutions. Our algorithm is compared with another algorithm that is representative of the state-of-the-art in evolutionary multiobjective optimization--NSGA-II. Simulation results on ZDT test problems show that ICMOA, in most problems, is able to find much better spread of solutions and better convergence near the true Pareto-optimal front compared to NSGA-II.

1 Introduction

Many real-world optimization problems involve optimization of several (conflicting) criteria. Since multiobjective optimization searches for an optimal vector, not just a single value, one solution often cannot be said to be better than another and there exists not only a single optimal solution, but a set of optimal solutions, called Pareto front. Consequently, there are two goals in multiobjective optimization: (i) to discover solutions as close to the Pareto front as possible, and (ii) to find solutions as diverse as possible in the obtained nondominate front. Satisfying these two goals is a challenging task for any algorithm for multiobjective optimization.

In recent years, many algorithms for mutiobjective optimization have been introduced. Most originate in the field of Evolutionary Algorithms (EAs). Among this the NSGA-II by Deb et al. [1] and SPEA2 by Zitzler et al. [2] are the most popular. IDCMA [3] designed for MO problems and CSADMO [4] used for dynamic MO problems originate in the field of Artificial Immune System.

The immune system is one of the most important biological mechanisms humans possess since our life depends on it. In recent years, several researchers have developed computational models of the immune system that attempt to capture some of their most remarkable features such as its self-organizing capability[5][6]. The immune system establishes the idea that the cells are selected when they recognize the antigens and proliferate. When exposed to antigens, the immune cells which may

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 100 – 109, 2006.

recognize and eliminate the antigens can be selected in the body and mount an effective response against them. Its main ideas lie in that the antigen can selectively react to the antibody, which are native production and spread on the cell surface in the form of peptides. The reaction leads to cell proliferate clonally and the colony has the same antibody. Some clonal cell divide into antibody producing cells, and others become immune memory cells to boost the second immune response. From the point view of the Artificial Intelligence, some biologic characters such as learning, memory and antibody diversity can be used to artificial immune system [7].

Based on the immune clonal theory, a new MO algorithm-Immune Clonal MO Algorithm (ICMOA) is proposed. We test ICMOA on the five ZDT test problems. The simulation results on the **five test problems** show that ICMOA outperforms the NSGA-II in terms of finding a diverse set of solutions and in converging near the true Pareto-optimal set.

2 The Artificial Immune System (AIS)

The main goal of the immune system is to protect the human body from the attack of foreign organisms. The immune system is capable of distinguishing between the normal components of our organism and the foreign material that can cause harm. These foreign organisms are called *antigens.* In AIS, antigen usually means the problem and its constraints. Especially, for the multiobjective optimization problems, we have

$$(P)\begin{cases}\min & F(\boldsymbol{x}) = (f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots f_p(\boldsymbol{x}))^T \\ S.T. & g_i(\boldsymbol{x}) \le 0 \quad i = 1,2,\cdots m \\ & h_j(\boldsymbol{x}) = 0 \quad j = 1,2,\cdots, p\end{cases} \tag{1}$$

Where, $\boldsymbol{x} = (x_1, x_2, \ldots x_n)$ and $\boldsymbol{x}$ is the vector of decision variables. $p \ge 2$ and p is the number of the objectives. $g_i(\boldsymbol{x}) \le 0 (i = 1,2,\cdots m)$ defines m inequality constrains and $h_j(\boldsymbol{x}) = 0 (j = 1,2,\cdots, q)$ defines q equality constrains.

The molecules called *antibodies* play the main role on the immune system response. The immune system response is specific to a certain foreign organism (antigen). When an antigen is detected, those antibodies that best recognize an antigen will proliferate by cloning. This process is called *clonal selection principle* [5]. In the traditional AIS, the new cloned cells undergo high rate mutations or hypermutation in order to increase their receptor population (called repertoire). These mutations experienced by the clones are proportional to their affinity to be antigen [6].Our Algorithm in this paper is based on the AIS previously described. In ICMOA, the affinity of the nondominated solutions in the objective space is defined as the same value 1 and that of the dominated solutions is defined as the same value 0.

3 The Multiobjective Optimization (MO)

Multiobjective optimization also called multicriteria optimization, multiperformance or vector optimization. Formally, we can state the general MO problem as follows:

$$\begin{aligned} \min \quad & \boldsymbol{y} = \boldsymbol{F}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots f_p(\boldsymbol{x}))^T \\ s.t. \quad & \boldsymbol{x} = (x_1, \cdots, x_n) \in \boldsymbol{X} \subset R^n \\ & \boldsymbol{y} = (y_1, \cdots, y_m) \in \boldsymbol{Y} \subset R^m \end{aligned} \tag{2}$$

where $\boldsymbol{x}$ is called decision vector, $\boldsymbol{X}$ decision space, $\boldsymbol{y}$ objective vector and $\boldsymbol{Y}$ objective space; $\boldsymbol{F}$ defines the mapping function. The following four concepts are of importance:

1. **Pareto dominance:** A solution $\boldsymbol{x}^*$ is said *dominate (Pareto optimal)* another solution $\boldsymbol{x}$ (denoted $\boldsymbol{x}^* \succ \boldsymbol{x}$) iff:

$$\forall i \in \{1, \cdots, m\} : f_i(\boldsymbol{x}^*) \le f_i(\boldsymbol{x}) \wedge \left(\exists k \in \{1, \cdots, m\} : f_i(\boldsymbol{x}^*) < f_i(\boldsymbol{x})\right). \tag{3}$$

2. **Pareto optimal:** A solution $\boldsymbol{x}^*$ is said to be *nondominated (Pareto optimal)* iff:

$$\neg \exists \boldsymbol{x} \in \boldsymbol{X} : \boldsymbol{x} \succ \boldsymbol{x}^*. \tag{4}$$

Pareto optimal solutions are also termed non-inferior or efficient solutions; their corresponding vectors are termed nondominated. These solutions may have no clearly apparent relationship besides their membership in the Pareto optimal set.

3. **Pareto optimal set:** The set P_S of all Pareto optimal solutions is defined as:

$$P_S = \left\{ \boldsymbol{x}^* \mid \neg \exists \boldsymbol{x} \in \boldsymbol{X} : \boldsymbol{x} \succ \boldsymbol{x}^* \right\}. \tag{5}$$

4. **Pareto optimal front:** The set P_F of all objective function values corresponding to the solutions in P_S:

$$P_F = \left\{ \boldsymbol{F}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots f_p(\boldsymbol{x}))^{\mathrm{T}} \mid \boldsymbol{x} \in P_S \right\}. \tag{6}$$

4 The Immune Clonal MO Algorithm (ICMOA)

4.1 The Algorithm

Our algorithm is the following:

Step1: Give the population size N, *the* clonal size R, the termination generation $G_{\max}$, the initial iteration $it := 1$; Generate randomly the initial antibody population with the size N.

Step2: Determine for each antibody in the antibody population, whether it is Pareto dominated or not.

Step3: Create a number R of copies of all of the nondominated antibodies.

Step4: Assign a mutation rate (MR) to each clone and apply mutation rate MR to each clone. In our algorithm, we adopt the *nonuniform mutation.*

Step5: Determine for each antibody in the new antibody population, whether it is Pareto dominated or not.

Step6: Delete all of the dominated antibodies in the antibody population and the corresponding individuals in the Pareto optimal set.

Step7: If the size of the antibody population is lager than N , delete the antibodies in the antibody population based on the crowding distance [1] until the size is N .

Step8: $it := it+1$ and go back to step3 until $it > G_{\max}$.

4.2 Nonuniform Mutation

In ICMOA, we adopt the nonuniform mutation. Z.Michalewicz [8] first combined the effect of mutation operator with the evolutionary generation. At the early age of the evolutionary process, mutation operation has a comparatively larger range; but as the evolutionary progresses, it becomes smaller and smaller which has a fine-tuning effect on the evolutionary system. The details are described as following:

Let $s = (v_1, v_2, \cdots, v_n)$ is a parent solution, and $v_k \in [a_k, b_k]$ is selected to mutate. The resulting solution after mutation is

$$s' = (v_1, \cdots v_{k-1}, v_k', \cdots, v_n) \tag{7}$$

where,

$$v_k' = \begin{cases} v_k + \Delta(t, b_k - v_k), & \text{if } \operatorname{rnd}(2) = 0 \\ v_k - \Delta(t, b_k - v_k), & \text{if } \operatorname{rnd}(2) = 1 \end{cases} \tag{8}$$

and rnd(2) is the result of managing a stochastic positive integer m with module 2 and t represents the current generation. The range of function $\Delta(t, y)$ is $[0, y]$ and when t becomes larger, the probability of $\Delta(t, y)$ approaching 0 increases.

Function $\Delta(t, y)$ is expressed as following:

$$\Delta(t, y) = y \bullet (1 - r^{(1-t/T)^{\lambda}}) \tag{9}$$

where, $r \in [0,1]$ is a random number, T is the maximum generation and parameter λ determines the degree of the nonuniformity and tunes the region of search.

5 Simulation Results

5.1 Test Problems

We analyze the performance measures for evaluating our algorithm on five ZDT test problems (introduced in [9]) that were frequently used as benchmark problems in the literature [1, 10, 11]. These problems are described in detail in table 1.

Table 1. Description of the Test Problems ZDT1, ZDT2, ZDT3, ZDT4 and ZDT6

ZDT1	
Decision space	$x \in [0,1]^{30}$
Objective functions	$f_1(x) = x_1$
	$f_2(x) = g(x)\left(1 - \sqrt{x_1 / g(x)}\right)$
	$g(x) = 1 + 9\left(\sum_{i=2}^{n} x_i\right)/(n-1)$
Optimal solutions	$0 \le x_1^* \le 1$ and $x_i^* = 0$ for $i = 2,\cdots,30$
Characteristics	convex Pareto front

ZDT2	
Decision space	$x \in [0,1]^{30}$
Objective functions	$f_1(x) = x_1$
	$f_2(x) = g(x)\left(1 - \left(x_1 / g(x)\right)^2\right)$
	$g(x) = 1 + 9\left(\sum_{i=2}^{n} x_i\right)/(n-1)$
Optimal solutions	$0 \le x_1^* \le 1$ and $x_i^* = 0$ for $i = 2,\cdots,30$
Characteristics	nonconvex Pareto front

ZDT3	
Decision space	$x \in [0,1]^{30}$
Objective functions	$f_1(x) = x_1$
	$f_2(x) = g(x)\left(1 - \sqrt{x_1 / g(x)} - x_1 / g(x)\sin(10\pi x_1)\right)$
	$g(x) = 1 + 9/(n-1)\sum_{i=2}^{n} x_i$
Optimal solutions	$0 \le x_1^* \le 1$ and $x_i^* = 0$ for
Characteristics	discontinuous Pareto front

ZDT4	
Decision space	$x \in [0,1]\times[-5,5]^9$
Objective functions	$f_1(x) = x_1$
	$f_2(x) = g(x)\left(1-\sqrt{x_1 / g(x)}\right)$
	$g(x) = 1+10(n-1)+\sum_{i=2}^{n}\left(x_i^2 - 10\cos(4\pi x_i)\right)$
Optimal solutions	$0 \le x_1^* \le 1$ and $x_i^* = 0$ for $i = 2,\cdots,10$
Characteristics	many local Pareto fronts

ZDT6	
Decision space	$x \in [0,1]^{10}$
Objective functions	$f_1(x) = 1-\exp(-4x_1)\sin^6(6\pi x_1)$
	$f_2(x) = g(x)\left(1-\left(f_1(x)/g(x)\right)^2\right)$
	$g(x) = 1+9\left(\sum_{i=2}^{n} x_i/(n-1)\right)^{0.25}$
Optimal solutions	$0 \le x_1^* \le 1$ and $x_i^* = 0$ for $i = 2,\cdots,10$
Characteristics	low density solutions near Pareto front

5.2 Performance Measures

Unlike in single-objective optimization, there are two goals in a multiobjective optimization: 1) convergence to the Pareto-optimal set and 2) maintenance of diversity in solutions of the Pareto-optimal set. In this paper, we adopt two performance metrics (proposed in [1]) that are direct in evaluating each of the above two goals in a solution set obtained by multiobjective optimization algorithm.

The first metric γ measures the extent of convergence to a known set of Pareto-optimal solutions. But this metric cannot be used for any arbitrary problem. It is suit for the test problems having a known set of Pareto-optimal solutions. The detail description of this metric please see[1].

The second metric Δ measures the extent of spread achieved among the obtained solutions. We first calculate the extreme solutions (in the objective space) by fitting a curve parallel to that of the true Pareto-optimal front. We use the following metric Δ to calculate the nonuniformity in the distribution:

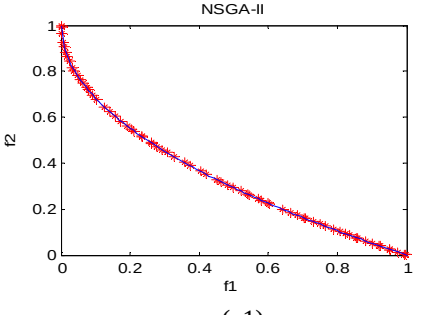
$$\Delta = \frac{d_f + d_l + \sum_{i=1}^{N-1}\left|d_i - \overline{d}\right|}{d_f + d_l + (N-1)\overline{d}} \tag{10}$$

Where, the parameters d_f and d_l are the Euclidean distances between the extreme solutions and the boundary solutions of the obtained nondominated set. The parameter d_i is the Euclidean distance between consecutive solutions in the obtained

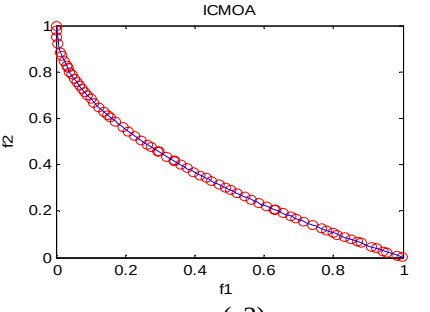
nondominated set of solutions and the parameter $\bar{d}$ is the average of all distance d_i, i=1,2$\cdots$,N-1 .

5.3 Discussion of the Results

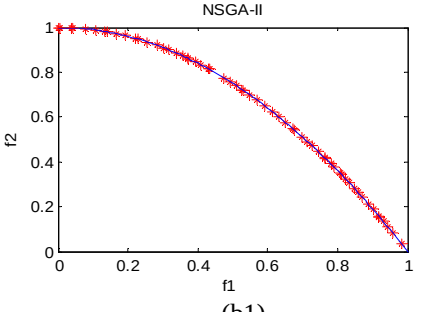
In order to validate our approach, we used five ZDT test problems that were frequently used as benchmark problems in the literature [1]. In the objective space, the known Pareto-optimal fronts are shown as continuous lines in NSGA-II and in ICMOA. The obtained solutions are shown with "*" in NSGA-II and with "o" in the ICMOA. The simulation results of NSGA-II and ICMOA on ZDT1, ZDT2, ZDT3, ZDT4 and ZDT6 are shown in figure 1.



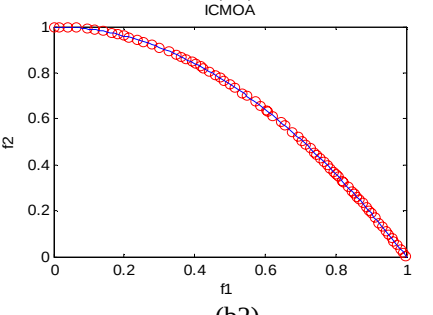
(a1)



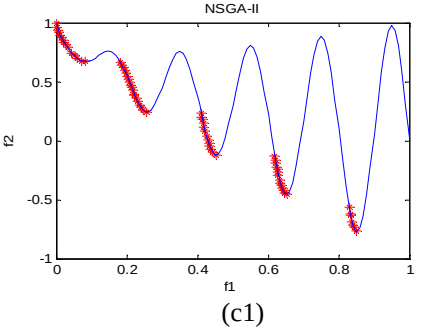
(a2)



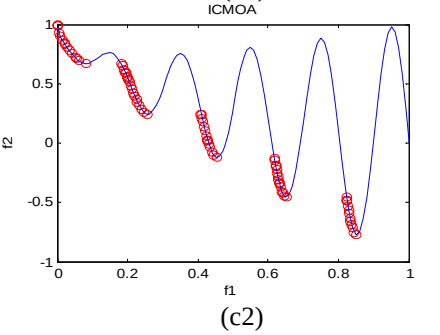
(b1)



(b2)



(c1)



(c2)

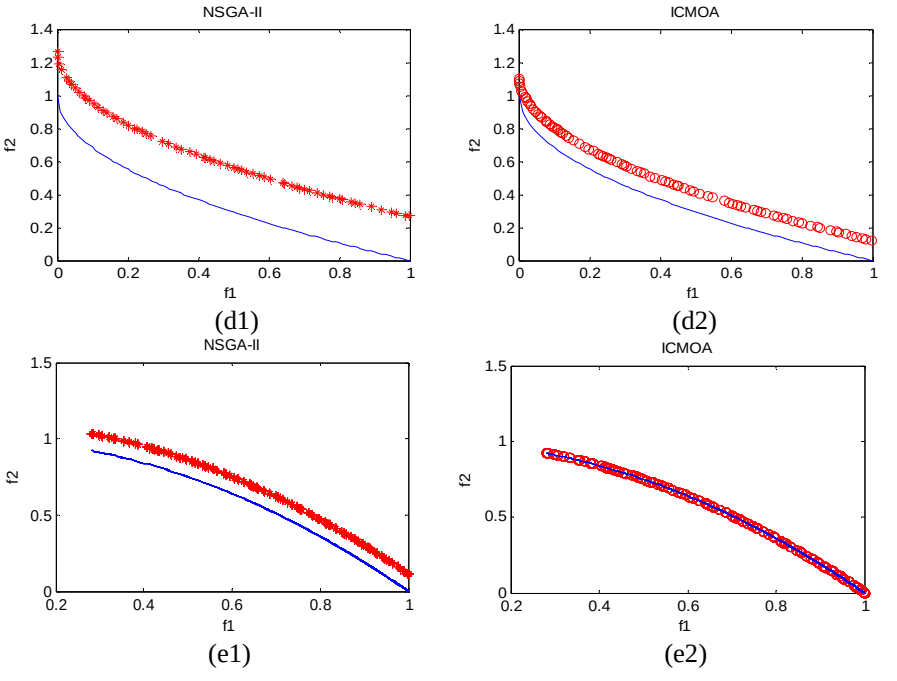


Fig. 1. Simulation results of NSGA-II and ICMOA on five ZDT test problems. (a1), (b1), (c1), (d1) and (e1) are the simulation results of NSGA-II on ZDT1, ZDT2, ZDT3, ZDT4 and ZDT6 respectively. (a2), (b2), (c2), (d2) and (e2) are the simulation results of ICMOA on ZDT1, ZDT2, ZDT3, ZDT4 and ZDT6 respectively.

Figure 1 shows that for ZDT1, ZDT2 and ZDT3, both NSGA-II and ICMOA have the ability in converging to the true front. But ICMOA has better ability in finding diverse solutions in the front. For ZDT4, NSGA-II and ICMOA encounter some difficulties in converging to the true front, but the convergence and the ability to find a diverse set of solutions much better with ICMOA. For ZDT6, the distribution in solutions is good with NSGA-II and ICMOA. ICMOA finds a definitely better converged set of nondominated solutions in ZDT6. However, NSGA-II has some difficulties in converging to the true front. So for this five ZDT test problems, our algorithm has a very competitive behavior with respect to NSGA-II.

Table 2. Mean and Variance of the Convergence Metric γ

Algorithm		ZDT1	ZDT2	ZDT3	ZDT4	ZDT6
ICMOA	Mean	0.0010	**0.0007**	**0.0129**	**0.2519**	**0.0040**
	Variance	0.0019	0.0005	0.0003	**0.0178**	0.0041
NSGA-II Real-coded	Mean	0.0335	0.0723	0.1145	0.5131	0.2966
	Variance	0.0048	0.0317	0.0079	0.1185	0.0131
NSGA-II Binary-coded	Mean	**0.0009**	0.0008	0.0434	3.2276	7.8068
	Variance	**0.0000**	**0.0000**	**0.0001**	7.3076	**0.0017**

Table 2 shows the mean and variance of the convergence metric γ obtained by ICMOA, NSGA-II (real-coded), NSGA-II (binary-coded). ICMOA is able to converge better in all problems except in ZDT1, where NSGA-II (binary-coded) found better convergence. Especially for ZDT4 and ZDT6, the mean of the convergence metric γ obtained by ICMOA is far smaller than NSGA-II (real-coded or binary-coded). In all cases with NSGA-II the variance in ten runs is small, except in ZDT4. In ZDT4, the variance obtained by ICMOA is much smaller than NSGA-II. This metric shows that ICMOA has better performance in the convergence to the Pareto-optimal front than NSGA-II.

Table 3. Mean and Variance of the Diversity Metric Δ

Algorithm		ZDT1	ZDT2	ZDT3	ZDT4	ZDT6
ICMOA	Mean	0.3995	**0.3950**	**0.3525**	**0.3327**	**0.2776**
	Variance	**0.0001**	**0.0013**	**0.0002**	**0.0010**	**0.0000**
NSGA-II	Mean	**0.3903**	0.4308	0.7385	0.7026	0.6680
Real-coded	Variance	0.0019	0.0047	0.0197	0.0646	0.0099
NSGA-II	Mean	0.4633	0.4351	0.5756	0.4795	0.6445
Binary-coded	Variance	0.0416	0.0246	0.0051	0.0098	0.0350

Table 3 shows that the mean and variance of the diversity metric Δ obtained by ICMOA, NSGA-II (real-coded), NSGA-II (binary-coded). ICMOA is able to find a better spread of solutions than NSGA-II (binary-coded or real coded) in all problems except in ZDT1, where NSGA-II (real-coded) found better spread of solutions. In all cases with ICMOA the variance in ten runs is small. This metric shows that ICMOA has better performance in finding the spread of solutions.

6 Conclusion

In this paper, we have introduced a new multiobjective optimization algorithm- Immune Clonal MO Algorithm (ICMOA). And test the algorithm on **five ZDT test problems which were** frequently used as benchmark problems. The ICMOA seems promising and is able to produce results similar to or better than those generated by an algorithm that represents the state-of-the-art in evolutionary multiobjective optimization. However, how to improve our algorithm to constrained test functions, how to produce a highly competitive algorithm (based on AIS) and how to evaluate a multiobjective optimization algorithm more efficiently are our further work and such work is currently under way.

References

1. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6(2002) 182-197.
2. Zitzler, E., Laumanns , M., Thiele. L.:SPEA2: Improving the Strength Pareto Evolutionary Algorithm. Technical Report 103, Computer Engineering and Networks Laboratory (TIK), Swiss Federal Institute of Technology (ETH) Zurich, Gloriastrasse 35, CH-8092 Zurich, Switzerland (2001)

3. Jiao, L.C., Gong, M.G., Shang, R.H., Du, H.F., Lu, B.: Clonal Selection with Immune Dominance and Anergy Based Multiobjective Optimization. Proceedings of the Third International Conference on Evolutionary Multi-Criterion Optimization, EMO 2005, Guanajuato, Mexico, March 9-11, 2005. Springer-Verlag, LNCS 3410, (2005) 474 – 489.
4. Shang, R.H., Jiao, L.C., Gong, M.G., Lu, B..: Clonal Selection Algorithm for Dynamic Multiobjective Optimization. Proceedings of the International Conference on Computational Intelligence and Security, CIS 2005. Springer-Verlag, LNAI 3801, (2005) 846 – 851.
5. de Castro, L. N., Von Zuben, F. J.: Learning and Optimization Using the Clonal Selection Principle. IEEE Transactions on Evolutionary Computation, Vol.6, No. 3, 2002: 239-251.
6. Coello, Coello, C., A. and Nareli, C. C.: An Approach to Solve Multiobjective Optimization Problems Based on an Artificial Immune System. In: Jonathan Timmis and Peter J. Bentley (editors). Proceedings of the First International Conference on Artificial Immune Systems (2002) 212-221
7. Du, H.F., Jiao, L.C., Wang, S.A.: Clonal Operator and Antibody Clone Algorithms. Proceedings of the First International Conference on Machine Learning and Cybernetics, Beijing, (2002) 506–510
8. Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Program. Springer-Verlag, Berlin. 1992
9. Zitzler, E., Deb, K., Thiele, L.: Comparison of multiobjective evolutionary algorithms: Empirical results. Evolutionary Computation 8 (2000) 173-195
10. Madavan,N.K.: Multiobjective optimization using a Pareto differential evolution approach. In: Congress on Evolutionary Computation (CEC'2002). Volume 2, Piscataway, New Jersey, IEEE Service Center (2002) 1145-1150
11. Xue, F., Sanderson, A.C., Graves, R.J.: Pareto-based multi-objective differential evolution. In: Proceedings of the 2003 Congress on Evolutionary Computation (CEC'2003). Volume 2, Canberra, Australia, IEEE Press (2003) 862-869
12. Licheng Jiao, Lei Wang. A novel genetic algorithm based on immunity. IEEE Transactions on Systems, Man and Cybernetics, Part A. Vol.30, No.5, Sept. 2000
13. L. Jiao, J. Liu, and W. Zhong, "An organizational coevolutionary algorithm for classification," IEEE Trans. Evol. Comput., vol. 10, no. 1, pp.67-80, Feb. 2006
14. J. Liu, W. Zhong, and L. Jiao, "A multiagent evolutionary algorithm for constraint satisfaction problems," IEEE Trans. Syst., Man, and Cybern. B, vol. 36, no. 1, pp. 54-73, Feb. 2006

Family Gene Based Grid Trust Model

Tiefang Wang, Tao Li, Xun Gong, Jin Yang, Xiaoqin Hu,
Diangang Wang, and Hui Zhao

School of Computer Science, Sichuan University, Chengdu 610065, China
wangtiefang@163.com

Abstract. This paper analyzes the deficiencies of current grid trust systems based on PKI (Public Key Infrastructure), ambiguity certificate principal information, and complicated identification process. Inspired by biologic gene technique, we propose a novel grid trust model based on Family Gene (FG) model. The model answers the existing questions in tradition trust model by adopting the technology of family gene. The concepts and formal definitions of Family Gene in the grid trust security domains are given. Then, the mathematical models of Family Gene are established. Our theoretical analysis and experimental results show that the model is a good solution to grid trust domain.

1 Introduction

Grid[1][2][3] is a new (next-generation) generation Internet. Current several grid trust models mainly are based on PKI theoretical [4][5]. They do not solve problem completely. Firstly, in this model, the legitimacy of the user node is guaranteed through the certificate that the CA distributes, but it is hard to differentiate homonymy entities in the reality due to ambiguity certificate principal information. Secondly, as identification process needs the third organization, the trust technique realization process is complicated. In the end, this technique needs great calculation capacity, and the efficiency is low. The gene technique was proposed by Gregor Mendel in 1865 abroad success application in the biology situation [6][7][8][9][10] enlightens us, leading it into the grid trust realm to have a kind of original creative meaning. This paper bring forward a new grid trust technology based on family gene principles and particular depict its application in grid trust domain.

2 Grid Trust Model Based on Family Gene

2.1 Grid Trust Model Based on Family Gene

We propose a grid trust model based on family gene and depict its workflow in detailed. The section ① in the figure 1 represents grid resource. The section ② in the figure 1 represents grid family based on family gene. The section ③ in the figure 1 represents gene examination and gene assignment. The work course of the model is as follows: The user of the application service is lined up one queue, gene-examining

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 110 – 113, 2006.

agents examine family gene of the first user in the sequential queue. If the user's family gene belongs to the family, an assigning gene is assigned the user. If the user's application conforms to the user privilege and behavior, user application will be satisfied; otherwise user's application is refused. If the user's family gene does not be-longs to the family, examining agent examine whether the user brings forward a regis-ter application or not. If the user puts forward the application, a new node is created for it and the user is endowed with the identity, otherwise user's application is refused.

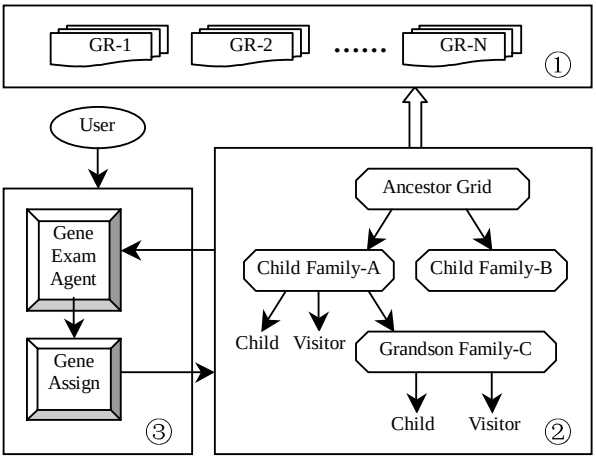


Fig. 1. The proposed grid trust model based on family gene

2.2 Formation Definition of Trust Model

According to grid trust model based on family gene, we describe a trust model for Grid family system in a structure (G, TR, OP), where G represents the set of the Grid family member gene, TR represents the set of trust relationships between Grid family member entities and OP represents the set of operations for the management of such trust relationships.

1) Gene of grid family (G)

Definition 1(Grid Family Gene): Grid family gene is a set of isogenous genes that have similar functions. The ancestral grid gene generated grid family gene by mutating and duplicating.

Definition 2(Aberrance Gene): Gene mutation generates the aberrance genes, which embody the differentiation of junior family members.

Definition 3(Assigning Gene): Assigning Gene is to assign a gene role (temporary visitor or Shaikh) to a user. The role is endowed with a certain privilege and behavior (read, write). In this model, the gene role is composed of two fields: privilege field, and behavior field. Users privilege and trust grade is ascertained according to users gene. The different privilege is endowed different behavior and function.

2) Trust Relationship between grid family members (TR)

Definition 4(Genetic Trust): Each children trust area is a subset of theirs father trust area. As long as the father-trust is satisfied, the sub-trust will be satisfied. The ancestor grid node has the largest trust field and privilege. So the user in the ancestor grid trust field can visit all children trust fields. The user in sub-trust field can only visit its children trust field, but can't visit the trust field of its father or ancestor. Therefore, the delivery of genetic trust is single direction, can't be reverse.

Definition 5(Brother Trust): Drawing lessons from the relation of brothers in the mankind, users of the brothers in the same class trust each other therefore they can visit each other, but the relation of the brothers' kid is delivery-trust at a certain time.

3) Operation in the grid trust model (OP)

Definition 6(Gene Identification): Gene's identification is to exam a visitor's gene to confirm whether it is the member of this family or not.

Definition 7(Gene Assignment): Gene assignment is to assign a gene role (temporary visitor or Shaikh) to a user. The gene role is endowed with a certain management behavior (Read, Write). In this model, we implemented gene assign behavior by assign an assigning gene to a user.

3 Simulations and Experiment Results

To test the validity of our approach, a series of simulation experiments of algorithm were performed. We developed a grid simulation toolkit based on GridSim [11] to satisfy our needs. The experiment was carried out in the Laboratory of Computer Network and Information Security at Sichuan University. In the imitate experiment, we compared FG grid trust model with PKI grid trust model of the time of examination user identity. The purpose of the experiment is to prove that the gene examination

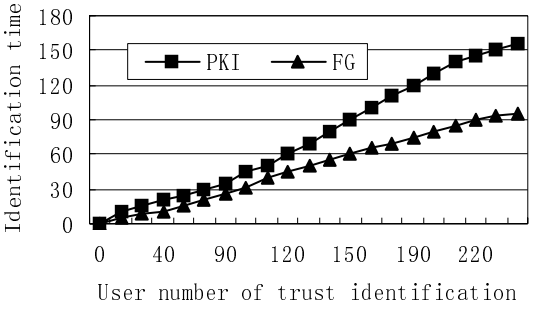


Fig. 2. The performance time and number of users. In the experiment we compared grid trust model based on Family Gene with grid trust model based on Public Key Infrastructure of the performance time and number of users.

method improves the efficiency of identity examination in the model. In the curve of figure 2, with the increasing of the user number, gene identification time in FG grid trust model increases more slowly than PKI. Its curve slowly goes up. But PKI curve goes up quickly. The method of FG model simplifies the process of the trust identification, hence reduces time of trust identification. It shows that the FG model avoids the overhead of trust identification calculation that improves the efficiency and performance of trust model.

4 Conclusions

Inspired by biologic gene technique, we carry out a beneficial attempt with gene technique application in the grid trust, present a new trust method different from a traditional PKI method, build up a grid trust model based on family gene, and give the formal definition of model and correlation conception of gene. We depict model work course and design a range of family gene trust algorithms and experiments. The experiment results prove the model has steady and excellent real-time performance on the realization and possess characteristic of simple and convenient identification.

References

1. R. B. K. Krauter and M. Maheswaran.: A taxonomy and survey of grid resource management systems software practice and experience. vol. 32, no. 2, p. 135164, Feb 2002.
2. C. K. I. Foster and S. Tuecke.: The anatomy of the grid: Enabling scalable virtual organizations. in International Journal of Supercomputer Applications, 2001.
3. D.W.Chadwick, A. Otenko, E.Ball.: Role-based Access Control with X.509 Attribute Certificates. IEEE Internet Computing, March-April 2003, pp. 62-69.
4. S. Tuecke, et al.: Internet X.509 Public Key Infrastructure (PKI) Proxy Certificate Profile (RFC 3820). IETF Network WG, June 2004. http://www.ietf.org/rfc/ rfc3820.txt.
5. G.L. Woods et al.: A mule cloned from fetal cells by nuclear transfer. Science, vol. 301, p. 1063, Aug. 2003.
6. T. Li.: Computer Immunology, Publishing House of Electronics Industry, Beijing, 2004.
7. T. Li.: An immune based dynamic intrusion detection model, Chinese Science Bulletin, vol. 50, pp. 2650 - 2657, November 2005.
8. T. Li.: An immunity based network security risk estimation, Science in China Ser. F Information Sciences, vol. 48, pp. 557-578, October 2005.
9. O. Mazda.: Improvement of nonviral gene therapy by Epstein-Barr virus (EBV)-based plasmid vectors. Curr Gene Ther, vol. 2, pp. 379-392, 2002.
10. M. Murshed, R. Buyya, D. Abramson.: GridSim: A Grid Simulation Toolkit for Resource Management and Scheduling in Large-Scale Grid Computing Environments. In Proc. of the 17th IEEE International Symposium on Parallel and Distributed (2002).

Immune Clonal Strategies Based on Three Mutation Methods

Ruochen Liu, Li Chen, and Shuang Wang

Institute of Intelligent Information processing, Xidian University,
Xi' an, 710071, China
ruocheenliu@yahoo.com.cn

Abstract. Based on the clonal selection theory, the main mechanisms of clone are analyzed in this paper, a new immune operator, Clonal Operator, inspired by the Immune System is discussed firstly. Based on the Clonal operator, we propose Immune Clonal Strategy Algorithm (ICSA); three different mutation mechanisms including Gaussian mutation, Cauthy mutation and Mean mutation are used in IMSA. IMSA based on these three methods are compared with Classical Evolutionary Strategy (CES) on a set of benchmark functions, the numerical results show that ICSA is capable of avoiding prematurity, increasing the converging speed and keeping the variety of solution. Additionally, we present a general evaluation of the complexity of ICSA.

1 Introduction

Biologically–motivated information processing systems can be classified into some branches such as Artificial Neural Network (ANN), Evolutionary Computation (EC) and Artificial Immune System (AIS). The great advances have been achieved in the fields of ANN and EC [1-3]. AIS is a relatively new bio-inspired area and aims at using ideas gleaned from immune system in order to develop systems capable of performing a wide range of tasks in various areas of research [4].

The models based on immune system principles, such as the clonal selection theory [5, 6], the immune network model [7, 8] or the negative selection algorithm[9], have been finding increasing applications in fields of science and engineering, such as: computer security, virus detection, process monitoring, fault diagnosis, pattern recognition, etc [10, 11].

According to Burnet [5], biological clonal selection occurs to the degree that a B-cell matches an antigen. A strong match causes a B-cell to be cloned many times, and a weak match results in little cloning. These 'clones' are mutated from the original B-cell at a rate inversely proportional to the match strength: a strongly matching B-cell mutates little and thus retains its match to the antigen to a slightly greater or lesser extent; a weakly matching cell mutates much more [12].

Based on the clonal selection theory, this paper presents a novel artificial intelligent algorithm, named Immune Clonal Strategy Algorithm (ICSA). The new immune operator, Clonal Operator, inspired by the Immune System is discussed firstly. Three different mutation mechanisms are used in ICSA; they are Gauss mutation [13], Cauchy mutation [14]

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 114 – 121, 2006.

and Mean Mutation [15]. For the sake of distinguishing, we call these algorithms as Immunity Clonal Strategy Algorithm based on the Mean Mutation (ICSAMM), Immunity Clonal Strategy Algorithm based on Gaussian Mutation (ICSAGM) and Immunity Clonal Strategy Algorithm based on Cauchy mutation (ICSACM) respectively. ICSA was compared with Classical Evolutionary Strategy (CES) on a set of benchmark functions.

This paper is organized as follows. Section 2 defines the clonal operator, and a novel algorithm of Artificial Immune System, Immunity Clonal Strategy Algorithm is put forward. In Section 3, we use the theory of Markov chain to prove the convergence of ICSA and present a general evaluation of the complexity of IMSA. Section 4 demonstrates the special implementation of ICSA used to compare with the Classical Evolutionary Strategy (CES), the numerical experiment shows ICSA can enhance the diversity of the population and avoid the prematurity to some extent. Finally Section 5 gives some conclusions.

2 Clonal Operator and ICSA

The clonal selection theory [5] is used by the immune system to describe the basic features of an immune response to an antigenic stimulus; it establishes the idea that only those cells that recognize the antigens proliferate, thus being selected. Upon exposed to antigens, an immune cell with an appropriate receptors will divide. Each daughter cell will have a receptor with the same receptor with the same specificity. In other words, an antigen specific cell will undergo clonal expansion. As a result, there will be many more cells capable of reacting with antigen, sufficient to mount a strong response, Some of this clones will differentiate into effector cells, and other will remain and will be ready to give rise to effector cells upon re-exposure of the host to the same antigen. The main features of the clone selection theory are:

- The new cells (clone) are copies of their parents, subject to a mutation mechanism with high rates;
- Elimination of newly differentiated lymphocytes carrying self-reactive receptors;
- Proliferation and differentiation on contact of mature cells with antigens;
- The persistence of forbidden clones, resistant to early elimination by self-antigens, as the basis of the autoimmune disease.

In order to use for reference this mechanism in artificial intelligent research fields, we propose a new operator—immune clonal operator.

2.1 Clonal Operator

The clonal operator is an antibody random map induced by the affinity including three steps: clone, clonal mutation and clonal selection. The state transfer of antibody population is denoted as follows:

$$C_{MA}:\ A(k) \xrightarrow{clone} A'(k) \xrightarrow{clonalmutution} A''(k) \xrightarrow{clonalselection} A(k+1) \qquad (1)$$

here antibody, antigen, the affinity between antibody and antigen are similar to the definitions of the objective function and restrictive condition, the possible solution, match between solution and the fitting function in AIS respectively. According to the

affinity function f (*), a point $a_i = \{x_1, x_2, \cdots x_m\}, a_i(k) \in A(k)$ in the solution space will be divided into q_i different points $a_i'(k) \in A'(k)$, by using clonal operator, a new antibody population is attained after performing clonal mutation and clonal selection.

If the antibody population is an n-dimension vector $A = \{a_1, a_2 \cdots a_n\}$, clone operator can be described as follows:

Clone T_c^C **:** The clonal operator is defined as

$$T_c^C(A) = [T_c^C(a_1) \quad T_c^C(a_2) \quad \cdots \quad T_c^C(a_n)]^T \tag{2}$$

Where $T_c^C(a_i) = I_i \times a_i \quad i = 1, 2 \cdots n$, and I_i is q_i dimension row vectors

$$q_i = g(N_c, f(a_i)) \tag{3}$$

Generally, q_i is given by:

$$q_i = N_c * \frac{f(a_i)}{\sum_{j=1}^{n} f(a_j)}, \quad i = 1, 2 \cdots n. \tag{4}$$

$N_c > n$ is a given integer relating to the clonal scale, it is easy to see that the clonal scale is regularized automatically. After cloning, the antibody population becoming the following equation:

$$A' = \{A, A_1', A_2', \cdots, A_n'\} \tag{5}$$

Where:

$$A_i' = \{a_{i1}, a_{i2}, \cdots, a_{iq_i - 1}\}, \quad a_{ij} = a_i, \quad j = 1, 2, \cdots, q_i - 1. \tag{6}$$

Clonal Mutation T_m^C **:** unlike the general mutation operator in GA, in order to save the information of the original population, the clonal mutation is unused to $A \in A'$ namely, the component x_i of individual a can be mutated as follows:

$$\eta_i'(j) = \eta_i(j) \exp(\tau' N(0,1) + \tau N_i(0,1)) \tag{7}$$

$$x_i'(j) = x_i(j) + \eta_i'(j)\delta \tag{8}$$

Where $x_i(j), x_i'(j), \eta_i(j), \eta_i'(j)$ denote the j^{th} component of the vector $x_i, x_i', \eta_i, \eta_i'$ respectively, $N_i(0,1)$ denotes a normally distributed one-dimensional random variable with mean zero and standard deviation one, τ and τ' are set commonly to $(\sqrt{2\sqrt{n}})^{-1}$ and $(\sqrt{2n})^{-1}$ [12], σ is a Gaussian distributed random variable or a Cauthy distributed random variable, or the combination between two distributed random variable.

Clonal Selection T_s^C **:** $\forall i = 1, 2, \cdots n$, if $b = \max\{f(a_{ij}) \mid j = 2, 3, \cdots q_i - 1\}$, let:

$$f(a_i) < f(b), \quad a_i \in A. \tag{9}$$

Then b replaces the antibody a_i in the original population. So the antibody population is updated, and the information communication between generations can be realized.

It is easy to find that the essential of the clonal operator is producing a variation population around the parents according to their affinity. Then the searching area is

enlarged, ulteriorly, the clonal operator maps a problem in a low dimension space (n dimensions) to a high one (N_c dimensions), and then projects the results to the original space after solving. Thereby the problem can be solved better.

2.2 Immune Clonal Strategy Algorithm

The main step of the ICSA is described as follows:

$k = 0$

Initialize the antibody population: $A(0) = \{a_1(0), a_2(0), \cdots, a_n(0)\} \in I^m$

Calculate the affinity: $A(0) : \{f(A(0))\} = \{f(a_1(0)), f(a_2(0)), \cdots f(a_n(0))\}$

While there is no satisfied candidate solution in $A(k)$, do

Clone: $A'(k) = T_c^C(A(k)) = [T_c^C(a_1(k)), T_c^C(a_2(k)) \cdots, T_c^C(a_n(k))]^T$

Clonal mutation: $A''(k) = T_m^C(A'(k))$

Calculate the affinity: $A''(k) : \{f(A''(k))\}$

Clonal selection: $A(k+1) = T_s^C(A''(k))$

Calculate the affinity:

$A(k+1) : \{f(A(k+1))\} = \{f(a_1(k+1)), f(a_2(k+1)) \cdots, f(a_n(k+1))\}$

$k = k + 1$

End

3 Analysis of the Proposed Algorithm

In this selection, we give a proof that the ICSA is convergent with probability 1and present a general evaluation of the complexity of ICSA.

3.1 Convergence of the Proposed Algorithm

Theorem 1: The population series of the immune clonal strategies algorithms $\{A(n), \quad n \geq 0\}$ is finite homogeneous Markov chain.

Proof: Similar to the evolutionary strategies**,** the state transfer of immune clonal strategy algorithms are processed on the finite space, therefore, population is finite, since

$$A(k+1) = T(A(k)) = T_s^C \circ T_m^C \circ T_c^C(A(k)) \tag{10}$$

T_c^C, T_s^C, T_m^C have no relation with *n*, so *A(n+1)* just relates with *A(n)*, namely, $\{A(n), \quad n \geq 0\}$ is finite homogeneous Markov chain.

The size of the initial population of the algorithms is *n* and the size of the middle population is N_c, all the possible solutions in the initial population can be seen as a point in the state space $S^1 := X^n$ and all the possible solutions in the middle population can be seen as a point in the state space $S^2 := X^{N_c}$, when it is not necessary to distinguish S^1 and S^2, we denote the state space as S, $s_i \in S$ denote the i^{th} state in S, let f is the affinity function defined on the hunting space X. Let:

$$s^* = \{x \in X \mid f(x) = \max_{x_i \in X} f(X)\} \tag{11}$$

Then we can define the convergence of the algorithms as follows.

Definition 1: Suppose for arbitrary initial distribution, the following equation satisfies:

$$\lim_{k \to \infty} \sum_{s_i \cap s^* \neq \varnothing} p\{A_k^i\} = 1 \tag{12}$$

Then we call the algorithm is convergent [16].

Theorem 2: The immune mono-clone strategies algorithm is convergent with the probability 1.

The definition of the convergence means that if the algorithm implements for enough iteration, then the probability with which the population contains the optimal individual will verge on 1, thus the definition shown as the above is usually called the convergence with probability 1. The detailed proof can be found in literature [17].

3.2 Analysis of the Complexity of the Proposed Algorithm

In this section we will present a general evaluation of the complexity of the proposed algorithm, taking into account the computational cost per generation for the real-valued coding.

We can measure the time required by an algorithm by counting the maximum number of instructions executed, which is proportional to the maximum number of times each loop ids executed. In the process of analysis, we will use the parameters that characterize the computations performed, such as the size of the antibody population n, the clonal size N_c; L is the number of variants to be optimized, in order to describe simply.

The proposed algorithm has four main processing steps: (1) determining the affinity of the antibodies, (2) clonal selection, and (3) the clonal mutation. According to [18], the step (1) and step (2) can be performed in $o(n)$ time for the worst cases. Mutating the individuals in the antibody population demands a computational time of the order $o(N_c \times L)$. The total computational time of algorithm will be:

$$o(n) + o(N_c \times L)$$

4 Numerical Experiment and Results

In order to validate the proposed algorithm, the ICSA is applied to solve optimization problem, this work also contains a compare done among CES, ICSAGM, ICSACM and ICSAMM based on the results of experiments. The following test functions are used to test the performance of the algorithms, all of them have more than one local optimum, and some even have infinite local optima solutions.

$$f_1(x) = x^2 + y^2 - 0.3\cos 3\pi x + 0.3\cos 4\pi y + 0.3 \; x, y \in [-1,1] \tag{13}$$

$$f_2(x) = 4x^2 - 2.1x^4 + \frac{1}{3}x^6 + xy - 4y^2 + 4y^4 \;\; x, y \in [-5,5] \tag{14}$$

$$f_3(x) = \sum_{i=1}^{10} x_i^2 \quad x_i \in [-100,100] \tag{15}$$

$$f_4(x) = \sum_{i=1}^{10} [x_i^2 - 10\cos(2\pi x_i) + 10] \; x_i \in [-5.12,5.12] \tag{16}$$

$$f_5 = \sum_{i=1}^{10} \left(\lfloor x_i + 0.5 \rfloor \right)^2 \quad x_i \in [-100,100] \tag{17}$$

$$f_6 = -20\exp\left(-0.2\sqrt{\frac{1}{10}\sum_{i=1}^{10} x_i} \right) - \exp\left(\frac{1}{10}\sum_{i=1}^{10} \cos(2\pi x_i) \right) + 20 + e \; x_i \in [-32,32] \tag{18}$$

For CES, the initial population size was 50. For ICSA, the antibody population size was 25 and the clone scale was 50. Given the optimized accuracy 10^{-3}, Table 1 provides a set of results averaged over 20 runs for the functions being optimized. The result records the mean numbers of iterations performed until the algorithms are capable of locating at least one global optimal. It can be sent from table 1ICSA outperform CES in 20 runs. It is interesting to note that ICSAMM, ICSAGM and ICSACM generally have a very similar convergence rate for all of test functions, and the convergent rate of ICSACM is higher than that of IMSAMM and IMSAGM when optimizing functions f_1-f_5, ICSAMM is performed extremely well in f_6. Additionally, it is demonstrated that bigger clonal size can improve the diversity; as a result, the prematurity can be avoided effectively.

Table 1. The performance comparison of ICSA and CES

Test function	Mean number of iterations and the standard deviation			
	CES	ICSACM	ICSAGM	ICSAMM
f_1	46 ± 14.87	25 ± 8.4	30.9 ± 7.8	28.9 ± 10.12
f_2	98 ± 20	27.3 ± 10.88	27.88 ± 13.91	28.5 ± 5.62
f_3	135 ± 24.2	85 ± 37.15	115.2 ± 25.56	110.16 ± 25.4
f_4	446 ± 94	62 ± 20.8	70 ± 11.94	85 ± 21.4
f_5	109 ± 19.3	71.5 ± 23.81	85 ± 37.1573	76.37 ± 23.81
f_6	198 ± 116.5	178 ± 84.2	98.5 ± 11.26	94.3 ± 20.7

With a further analysis of the optimizing results, we find that ICSA can hunt more local optima than CES. Figure 1 shows a certain optimizing results of function f_1 by CES and ICSA, here "*"denotes the optimal value.

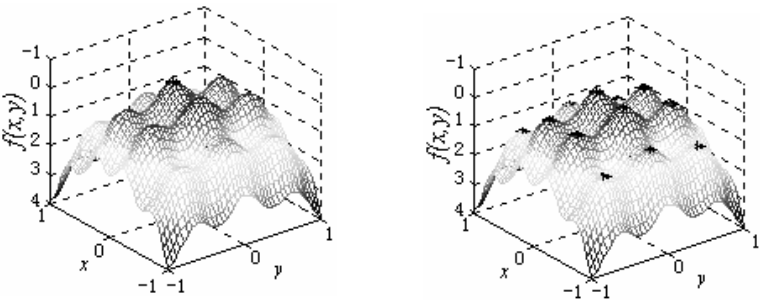


Fig. 1. The optimizing results of $f_1(x, y)$

5 Conclusions

The mechanize of the antibody clonal selection is discussed systematically in this paper; the clonal operator is proposed, by applying the clonal operator, Immune Clonal Strategy Algorithm is put forward. We find the essential of the clonal operator is producing a variation population around the parents according to their affinity, and then the searching area is enlarged. Compared with CES, ICSA is convergent faster and the diversity is much better. A further study shows when the objective function becomes complex and the coupling between variables are great, the hunting ability of the algorithms becomes weak, the later research will focus on improving the ability of the algorithms to hunt the local optima based on the ideas of intelligent-synthesizing.

References

[1] De Castro, L. N. and Von Zuben, F. J.: Artificial immune systems: Part I—basic theory and applications. Technical Report DCA-RT 01/99, School of Computing and Electrical Engineering, State University of Campinas, Brazil, 1999. Available: http://www.dca.fee.unicamp.br/~Inunes/immunes.html.

[2] Jiao, L. C, Liu, J. and Zhong, W. C.: An organizational coevolutionary algorithm for classification. IEEE Trans. Evol. Comput. 2006, 10(1): 67-80.

[3] Liu, J., Zhong, W. C. and Jiao, L, C.: A multivalent evolutionary algorithm for constraint satisfaction problems. IEEE Trans. Syst., Man, and Cybern. B, vol2006, 36(1): 54-73

[4] Forrest, S. and Steven, A. Hofmeyr.: Immunology as Information Processingfrom Design Principles for Immune System & Other Distributed Autonomous Systems. L. A. Segel and I. R. Cohen, eds. Oxford Univ. Press, 2000:361-387. www.cs.unm.edu/~forrest/ publications/iaip.pdf.

[5] Burnet, F. M.: The Clonal Selection Theory of Acquired Immunity. Cambridge University Press.1959.

[6] Forsdyke, D. R.: The origins of the clonal selection theory of immunity as a case study for evaluation in science. The FASEB Journal, 1995, 9:164-166.

[7] Jerne, N. K.: The natural-selection theory of antibody formation. Proceedings of the National Academy of Sciences USA, 1955, 41: 849-856.

[8] Coutinho, A.: The network theory: 21 years later. Scandinavian Journal of Immunology, 1995, 42: 3-8.
[9] Jerne, N.K.: Towards a network theory of the immune system. Ann. Immunology. (Inst. Pasteur), 1974, 125(C): 373-389.
[10] DasGupta, D.: Artficial Immune Systems and Their Applications. Springer-Verlag New York, Inc.1998.
[11] De Castro, L. N. and Timmis. J.: Artificial Immune Systems: A New Computational Intelligence Approach, Springer-Verlag, 2002.
[12] French, D. L., Reuven, L., and Scharff, M. D.: The role of somatic hypermutation in thegeneration of antibody diversity. Science, 1989, 244(4909): 1152-1157.
[13] Fogel, D. B., Atmar, J. W.: Comparing Genetic Operators with Gaussian Mutations in Simulated Evolutionary Processes Using Linear Systems. Biological Cybernetics, 1993, 63: 111-114.
[14] Yao, X., Liu, Y. and Lin, G.: Evolutionary Programming Made Faster, IEEE Transactions on Evolutionary Computation, 1999, 3(2): 82-102.
[15] Chellapilla, K.: Combining Mutation Operators in Evolutionary Programming, IEEE Transactions Computation, 1998, 2(3): 91-96.
[16] Jiao, L. C., Wang, L.: A Novel genetic Algorithm based on Immunity. IEEE Trans. On Systems, Man, and Cybernetics-Part A Systems and Humans, 2000, 30(5): 552-551.
[17] Liu, R. C., DU, H. F. and Jiao, L. C.: Immunity Ploy clonal Strategy. Journal of Computer Research and Development. 2004, 41(4): 571-576.
[18] Fischetti, M .and Martello, S.: A Hybrid Algorithm for finding the kth Smallest of n Elements in $O(n)$ time, Ann. Operation Res., 1998 ,13: 401-419.

A High Level Stigmergic Programming Language

Zachary Mason

Efficient Frontier Inc.
Mountain View, CA
zmason@efrontier.com

Abstract. Terrestrial social insects build architecturally complex nests despite their limited sensors, minimal individual intelligence and the lack of a central control system. [3] Many of the nest structures emerge as a response of the individual insects to pheremones, which the insects themselves can emit.[2] The work in [4] extrapolated from social insect building behavior to a system where the behavior of homogenous swarms of virtual agents could be designed to build simple structures. Like termites, these agents have no memory and limited sensors, and the macroscopic structure emerges from their interactions with their immediate environments. This paper presents Stigcode, a swarm programming language that permits more complex structures to be more conveniently specified. A StigCode program is a description of a target structure that is compiled into a set of reactions to pheremone concentrations for the swarm agents. Though not Turing-Universal[1], StigCode provides a syntax for defining re-usable, composable design elements. In keeping with the entomorphic theme, In the manner of ant and termite nests, StigCode architectures can do limited self-repair

Keywords: stigmergy, swarm intelligence, stigmergic programming, ant algorithms, self-organization.

1 Introduction

Termite nests are large, intricate structures, ten- to a hundred-thousand times the size of an individual termite. This structure-size-to-creator-size ratio is exceeded only by the largest human cities [1]. The African termite sub-family Macrotermitinae builds complex nests of many distinct functional parts, such as defensive walls, pillar-supported brood chambers, spiral cooling shafts, royal sanctuaries and fungus gardens. Although the termites' nests are architecturally advanced, the termites themselves are blind, physcially weak, have relatively simple nervous systems and seem to lack any central authority (the queen, despite her name, is essentially an inert egg-laying factory.)

[1] Since it is, in effect, a language for specifying architecture and from that specification producing instructions that will emergently produce it, recursion is unnecessary (unless, perhaps, the specified architecture is a sort of Borgesian architectural nightmare.)

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 122–125, 2006.

[4] describes a way to generalize the termite's approach to construction (as described in [3],[2],,[1]) and use it to develop articial, decentralized swarms that can be programmed to build complex structures, although the programming is onerous - akin to writing a program in assembly language. This work builds on that, providing a powerful, higher-level combinatorically defined language for specifying the structures a swarm is to build.

The principle underlying the communication of termites and many other social insects is *stigmergy*, that is, communication via changes in the environment rather than through the direct transmisssion of information. Stigmergy has been used to model the pillar-building behavior of termites, the foraging behavior of ants, and their spontaneous cemetery building. These applications are examples of qualitative stigmergy - the actions of individual agents are governed by continuous environmental variations.

2 The Command Language

StigProg is a command language for programming swarms of agents to stigmergically build structures. The virtual world of the agents is a two dimensional grid. Each cell of the grid contains one or zero agents, one or zero bricks (the construction material in the StigCode universe) and the some concentration of pheremones. The agents wander around at random - at each timestep, they can pick up a brick, put down a brick, or emit a pheremone plume. Which action they will undertake depends on the local pheremone concetrations. Pheremone intensities decrease geometrically over distance. Pheremones do not interact with the world except in so far as they stimulate the agents to various behaviors.

The swarm of agents is homogenous. Agents have no memory and no state - they react to their environments from a condition of innocence. They form a sort of gas, permeating the grid-space, enforcing the rules compiled from the StigCode at randomly selected points, thereby influencing the behavior of the other agents. Agent behavior is governed by a set of behaviors - a behavior is triggered if pheremone concentraitons at the agent's location fall within a specified range.

The limitations of agents and their dependence on their immediate environment for information suggest an interesting gestalt shift, namely that it is less useful to think of the StigCode as a way of programming with swarms, than as a way of programming with interacting pheremone plumes implemented via swarms. In effect, the agents are a lower level computational stratum underlying a cellular automaton over pheremone plumes.

The primary problem one encounters when programming with pheremone plumes is that the pheremone concentration diffuses isotropically away from a given deposit - that is, there is no directional orientation and no way to tell left from right (it would be unfortunate to be restricted to structures with mirror symmetries.) Also, concentrations of pheremone rapidly become very low as one gets farther from the deposit, making it difficult, if there is noise or a competing signal in the environment, to accurately assess the distance to the deposit.

These problems is solved by using multiple pheremone plumes to define lines and preserve orientation. Part of the challenge of designing StigCode was ensuring that this information was coherently maintainted at each step in the run of the program.

StigCode is something between a command language and an architectural specification language. A StigCode program indirectly describes the shape of the structure that the swarm is to build, specifying it in terms of the combinatorics of a set design primitives. The most elementary StigCode structure is a base-point, which consists of four pheremone plumes, required so that, based on pheremone concentration, both distance and orientation with respect to the base-point can be inferred.

Base-points are the stepping off points for other structures, such as line segments. A line segment emanates from a base-point at a given angle, for a given length. The line is built by incrementally adding pheremone deposits, each close to the last one laid down. This takes care of the problem of the difficulty of communication over long distances (as there may be other, similar lines in existence, remotely exerting an unwelcome influence), but introduces a tendency to drift, as, since behaviors are based on continuous plume gradients, there is a small range of points at which each successive plume may deposited, and errors may accumulate.

Each command in StigCode is compiled down to a lower-level command language that is intended to make it easy to program behaviors with respect to the geometry of pheremone deposits (i.e., in terms of distance and angle) rather than absolute local pheremone levels, as described in [4].

Below is a program used to encode an approximation of a chess board. The plan is essentially to build a sort of spinal column heading downward from the upper left hand corner. Periodically this column emits either an even or an odd row, heading east. The rows periodically emit pheremones that code for the construction of a square.

```
grow-segment:begin vertebra1,angle 0,end vertebra2,steps 8,name spine1
grow-segment:begin vertebra2,angle 0,end vertebra1,steps 8,name spine2
grow-segment:begin spine1,angle -90,end chess1,steps 4,name spine-row1
grow-segment:begin spine2,angle -90,end chess2,steps 4,name spine-row2
grow-segment:begin chess1,angle 0,end chess2,steps 4,name row1
grow-segment:begin chess2,angle 0,end chess1,steps 4,name row2
build-square:center chess1, radius 4, name makesquare, deposit true
```

In the system's initial state, both bricks are agents are distributed at random. Figure 1 shows a final state of the system - the bricks are all collected into a chess-board like pattern, though there aren't quite enough bricks to complete the pattern. The squares have hollow centers because an agent carrying a brick releases the brick as soon as it steps into the pheremone boundary for that behavior (and is on an empty square.)

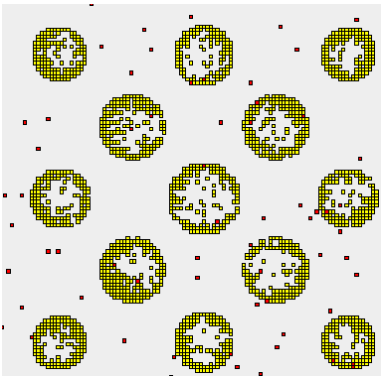


Fig. 1. End state of chess-board builder

3 Conclusion

StigCode is a programming language with re-usable, composable elements for specifying swarm construction. Generalizing on termites, the swarms consist of simple homogenous agents with limited sensors who communicated through virtual pheremone plumes.

References

1. [Bonabeau et al 97] Boneabeau, Theraulaz, Deneubourg, Franks, Rafelsberger, Joly, and Blaco: A model for the emergence of pillars, walls and royal chambers in termite nests Phil. Trans. R. Soc. London (1997)
2. [Bruinsma 79] Bruinsma, O. H: An analysis of building behaviour of the termite Macrotermes subhyalinus Ph.D. Dissertation, Landbouwhogeschool, (1979)
3. [Grasse 84] Grasse, P: Termitologia, Tome II - Fondation des societies construction. Paris: Masson. (1984)
4. [Mason 02] Z. Mason: Programming Stigmergy - Using Swarms for Construction Artificial Life VIII (2002)
5. [Werfel et al. 05] J. Werfel, Y. Bar-Yam and R. Nagpal: Building Patterned Structures with Robot Swarms. Proc. IJCAI (2005)

Application of ACO in Continuous Domain

Min Kong and Peng Tian

Shanghai Jiaotong University, Shanghai 200052, China
kongmin@sjtu.edu.cn,
ptian@sjtu.edu.cn

Abstract. The Ant Colony Optimization has gained great success in applications to combinatorial optimization problems, but few of them are proposed in the continuous domain. This paper proposes an ant algorithm, called Direct Ant Colony Optimization (DACO), for the function optimization problem in continuous domain. In DACO, artificial ants generate solutions according to a set of normal distribution, of which the characteristics are represented by pheromone modified by ants according to the previous search experience. Experimental results show the advantage of DACO over other ACO based algorithms for the function optimization problems of different characteristics.

Keywords: Ant Colony Optimization, Function Optimization Problem, Continuous Domain.

1 Introduction

Swarm Intelligence is a fast growing meta-heuristic of Evolutionary Computation [16,14]. Of which the Ant Colony Optimization (ACO) is one of the most promising benchmark applications, which is inspired from the stigmergy communication of the real ants [4,6].

Although ACO has gained great success in some combinatorial optimization problems [5,10,9], the application to the function optimization problem appears a little difficult, since the pheromone laying method is not straightforward, specially in the continuous domain. There have been several attempting applications of ant algorithm in the continuous domain, they are: the Continuous ACO(CACO) by Bilchev etal. [1,2,11,15], the API algorithm by Monmarchè [12], Continuous Interacting Ant Colony (CIAC) by Dréo and Siarry[7,8], and the extended ACO application, proposed by Socha[13].

CACO consists of global and local search, the global search incorporates a GA method, and the local search is a kind of ant algorithm. API incorporates a moving nest idea inspired by the prey behavior of the Pachycondyla apicalis ants. CIAC does not use the stigmergy communication, but uses a kind of direct communication within two communication channels. The extended ACO is much closer to the framework of ACO, it incorporates a mixture of Probability Density Function (PDF) to replace the discrete probability distribution. The common problem of these algorithms lies in their weak performance comparing to other state-of-art algorithms in continuous domain.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 126–135, 2006.

This paper proposes a Direct Ant Colony Optimization algorithm (DACO) for tackling the function optimization problem in continuous domain. Different from other ACO applications to the function optimization problems, pheromone in DACO are directly associated to a set of normal distribution. Artificial ants construct solutions by generating stochastically the value of every variable according to the normal distribution. DACO is developed on the base of the general ACO framework. It takes advantage of the successful results of recent ACO studies by the use of some successful ideas such as some special pheromone maintenance methods. Experimental results over various benchmark problems show the potential of DACO for solving function optimization problems of different characteristics.

2 ACO in Continuous Domain

Various versions of ACO have a common framework which is described in pseudocode in fig.1[6]. The main procedure of the ACO metaheuristic, ScheduleActivities, manages the scheduling of three components:

ConstructAntsSolutions A population of artificial ants are managed to construct incrementally the solutions to the optimization problem instance. They add solution components one by one with a stochastic local decision policy that make use of pheromone trails and problem specific information.

UpdatePheromones Pheromone trails are modified according to the previous search experience. Normally it includes two phases of procedure: the evaporation phase and the intensification phase. The former avoids a too rapid convergence of the algorithm toward a local optima, the latter focuses the search around the rather promising area.

DaemonActions This is the optional component which is used to implement centralized actions that cannot be performed by single ant, such as local search, pheromone re-initialization, and search process control etc.

The main difference of the various version of ACO is the pheromone laying methods and the way of handling the above three components, how to schedule and synchronize these components according to the characteristics of the considered problem is key to the quality of the algorithm.

DACO is designed under the ACO metaheuristic framework, in which an iteration includes the three components as described above. At the following subsections, we will describe in detail the ConstructAntSolutions and Update Pheromones procedure, the DaemonActions that contains more centralized control will be discussed in the next section, in which some previous successful implementation experience of different version of ACO, such as the methods of pheromone maintenance and the use of local search, are incorporated into the system.

```
Procedure ACOMetaheuristic
  ScheduleActivities
      ConstructAntsSolutions
      UpdatePheromones
      DaemonActions        % optional
  end-ScheduleActivities
end-procedure
```

Fig. 1. The ACO metaheuristic in pseudo-code

2.1 Solution Construction

In DACO, a number of artificial ants are managed in searching the continuous domain of the problem instance at each iteration. Each artificial ant constructs its solution by generating a real number for each variable i stochastically decided by a special normal distribution $N(\mu_i, \sigma_i^2)$, of which the characteristics are modified by the ants during iterations in a form of pheromone maintenance procedure.

For a normal distribution $N(\mu, \sigma^2)$, the mean value μ and the deviation value σ completely identify its characteristics, the probability density function (PDF) of a normal distribution $N(\mu, \sigma^2)$ is:

$$f_{PDF}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty \tag{1}$$

μ represents the center of the distribution, or the concentrated position of the random variables, the left part of fig.2 shows the positions of PDF with different μ and a fixed σ; σ represents the diversification of the distribution, with a small σ, the PDF will be narrow and high, with a quick descendence along both sides of $x = \mu$, representing that the random variables are more concentrated around $x = \mu$; while a large value of σ means that the PDF is flat, with a slow descendence along both sides of $x = \mu$, representing that the random variables are more diversified. The right part of fig.2 shows the PDF with different σ and a fixed μ.

DACO incorporates a number of n normal distributions, with each normal distribution $N(\mu_i, \sigma_i^2), i \in \{1, \cdots, n\}$ associated to each variable i, where n is the number of variables in the function optimization problem. And two kinds of

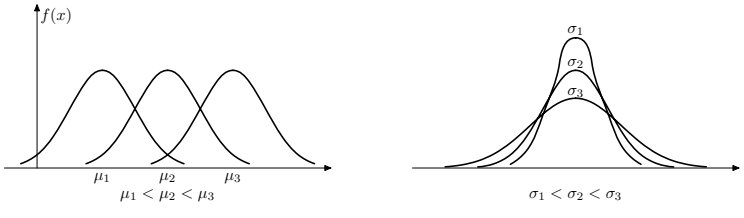


Fig. 2. Characteristics of the Normal Distribution

pheromone are incorporated: one is associated to $\mu = [\mu_1, \cdots, \mu_n]$, the other is associated to $\sigma = [\sigma_1, \cdots, \sigma_n]$. The amount of pheromone associated with μ and σ directly represent the value of μ and σ respectively. In the following, we will use the term μ and σ as the pheromone representation as well.

The solutions generated according to the normal distribution are checked and modified to fit into the constraint range, that is, for every variable x_i, it follows the following transferring rule:

$$x_i = \begin{cases} a_i & x_i < a_i \\ x_i & a_i \leq x_i \leq b_i \\ b_i & x_i > b_i \end{cases} \tag{2}$$

where b_i and a_i are the top and lower limit of variable x_i.

2.2 Pheromone Update

Initially, pheromone are set as $\mu_i(0) = a_i + rand()(b_i - a_i)$, and $\sigma_i(0) = (b_i - a_i)/2$, where $rand()$ is a randomly generated number with an even distribution in the range of $[0, 1]$.

After all the ants have finished their solution construction, the pheromone evaporation phase is performed first, in which all the pheromone evaporate :

$$\begin{aligned} \mu &\leftarrow (1-\rho)\mu \\ \sigma &\leftarrow (1-\rho)\sigma \end{aligned} \tag{3}$$

where $\rho \in [0, 1]$ is the pheromone evaporation parameter.

Then, the pheromone intensification procedure is performed as:

$$\begin{aligned} \mu &\leftarrow \mu + \rho x \\ \sigma &\leftarrow \sigma + \rho|x - \mu| \end{aligned} \tag{4}$$

where $x = [x_1, \cdots, x_n]$ is the solution used for pheromone intensification, which is normally the global best solution S^{gb} found by the previous search.

3 Additional Features

Theorem 1. *Suppose S_L is a local optima, and there has been no improvement over S_L after t iterations in DACO, we have $\mu \to S_L$ and $\sigma \to 0$ when $t \to \infty$.*

Proof. According to the pheromone update procedure described in the previous section, we can combine equation (3) and (4) to calculate the μ at iteration t as:

$$\begin{aligned} \mu(t) &= (1-\rho)\mu(t-1) + \rho S_L \\ &= (1-\rho)^t \mu(0) + \textstyle\sum_{i=1}^{t} (1-\rho)^{t-i} \rho S_L \end{aligned} \tag{5}$$

Because $0 < \rho < 1$, asymptotically, we have $\mu \to S_L$ when $t \to \infty$. For the same reason, we have,

$$\begin{aligned} \sigma(t) &= (1-\rho)\sigma(t-1) + \rho|S_L - \mu(t-1)| \\ &= (1-\rho)^t \sigma(0) + \textstyle\sum_{i=1}^{t} (1-\rho)^{t-i} \rho|S_L - \mu(i-1)| \end{aligned} \tag{6}$$

Because $\mu \to S_L$ with $i \to t \to \infty$, we have $\sigma \to 0$.

Theorem 1 means that once DACO gets into a local minima S_L, it will be more and more likely to find S_L again, in other words, DACO will get into premature very easily. The convergence problem of DACO is rather how to escape from a local optima than how to get into it.

DACO incorporates three methods in the DaemonActions procedure to escape from a local optima, they are, local search with different neighborhood structure, pheromone re-initialization, and using different solutions for pheromone intensification.

Local Search. In DACO, a simple local search is designed, in which a random bidirectional movement is performed one by one for every variable. The local search for every variable x_i contains two phases: the first phase increase x_i with a random distance d_i, and continues until a worse solution is found or the constraint limitation are violated. The second phase reverse the process by decrease x_i with a random distance d_i. The random distance d_i moved at every step is biased by σ_i

$$d_i = rand() * \sigma_i \tag{7}$$

The local search process is performed to the iteration best solution S^{ib} and the global best solution S^{gb} during every iteration cycle.

Pheromone Re-initialization. The main idea is to re-initialize all the pheromone once the algorithm gets near to a local optima. In DACO, a convergence factor, cf, is defined to monitor the status of convergence:

$$cf = \frac{\sum_{i=1}^{n} \frac{2\sigma_i}{b_i - a_i}}{n} \tag{8}$$

Initially, since σ_i is set as $\sigma_i = (b_i - a_i)/2$, thus that $cf = 1$; when the algorithm gets into some local optima, σ will be decreased to be close to 0 as indicated by theorem 1, such that $cf \to 0$. So, according to the definition of cf, we can see that as the algorithm gets near to a local optima, cf changes from 1 to 0, we can set a threshold cf_r as a trigger for the re-initialization, once $cf \leq cf_r$, the algorithm is considered to be in some state of premature, then a re-initialization procedure will be performed, in which the pheromone μ is set equal to S^{gb}, σ_i is set to the initial value $(b_i - a_i)/2$, followed directly by a pheromone intensification phase using all the previous re-initialized best solutions S^{rb}s as a weighted combination for pheromone intensification:

$$\begin{aligned} \mu &\leftarrow \mu + \rho \sum_{j=1}^{n_r} w_j S^{rb}(j) \\ \sigma &\leftarrow \sigma + \rho \sum_{j=1}^{n_r} w_j |S^{rb}(j) - \mu| \end{aligned} \tag{9}$$

where n_r is the number of previous S^{rb}, w_j is the weight for the jth S^{rb}, which is calculated according to the fitness value of every S^{rb}:

$$w_j = \frac{f(S^{rb}(j))}{\sum_{l=1}^{n_r} f(S^{rb}(l))} \tag{10}$$

Using Different Solutions for Pheromone Intensification. Using different solution for pheromone intensification is another method for diversification. We use the method introduced by Blum and Dorigo [3], where different combination of S^{ib}, S^{rb} and S^{gb} are selected for the pheromone intensification according to the current status of convergence. The details of the selecting strategy is described in Tab.1, where w_{ib}, w_{rb} and w_{gb} are the weights for the solution S^{ib}, S^{rb} and S^{gb} respectively. cf_is are the threshold parameters whose values will be determined in the next experimental section.

Table 1. Pheromone Intensification Strategy for DACO

	$cf > cf_1$	$cf \in [cf_2, cf_1)$	$cf \in [cf_3, cf_2)$	$cf \in [cf_4, cf_3)$	$cf \in [cf_r, cf_4)$
w_{ib}	1	2/3	1/3	0	0
w_{rb}	0	1/3	2/3	1	0
w_{gb}	0	0	0	0	1

4 Experimental Results

Two sets of experiments are performed in this paper. The purpose of the first set is to fine tune the parameter m, ρ and cf_i for DACO. The second test set is to compare DACO with other Ant System. Five mostly used test functions in the literature are selected as the test bed for comparison. These test functions are also used in the parameter setting experiments, the detail of these functions are presented in Appendix A. In all the experiments performed in this paper, a test is considered to be successful if it observes the following condition:

$$|f^* - f_{known\_best}| < \epsilon_1 \cdot |f_{known\_best}| + \epsilon_2 \tag{11}$$

where f^* is the optimum found by DACO, f_{known_best} is the global optimum of the test function. ϵ_1 and ϵ_2 are accuracy parameters, which is set to be: $\epsilon_1 = \epsilon_2 = 10^{-4}$.

4.1 Primary Experiments for Parameter Setting

As discussed in the previous section, there are three parameters need to be set by experiments, they are cf_i, m and ρ. We set the parameters one by one with other parameters being fixed, the algorithm stops when it satisfies condition (11).

Parameter Setting for the cf_i Threshold. To get a good cf_i setting, a set of 25 combinations of cf_i are designed for test considering a wide range of different combinations. Fig.3 shows the test results, which includes the accumulated best and average evaluation numbers for all the test functions over 100 runs.

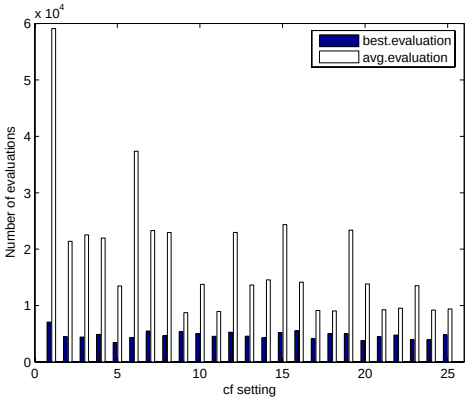


Fig. 3. Test Results of cf setting

From the test results we can see that several combinations of the cf_i settings have better performance than others, such as the combinations of 9, 11, 17, 18, 21, 22, 24 and 25. Among them, the combination 9 is the best, which is $cf = [1.0, 0.5, 0.1, 0.05, 0.01]$. In all the following tests, we will use this cf_i combination.

Parameter Setting for m. Parameter m is tested with a range from 1 to 20 for all the 5 test functions. Fig.4 shows the test results of average evaluation number over 100 runs for each test function. From the test results we can see that, overall, more m needs more evaluation number, but a number of 1 or 2 do not definitely get the best performance, a number of 3 or 4 yield the overall best results. For the following tests, we will use $m = 3$ as the parameter setting.

Parameter Setting for ρ. Parameter ρ is tested with a range from 0.01 to 1 with a step of 0.01. The accumulated average evaluation of the five test function is presented in fig.5. From fig.5 we can see that the average evaluation decreases

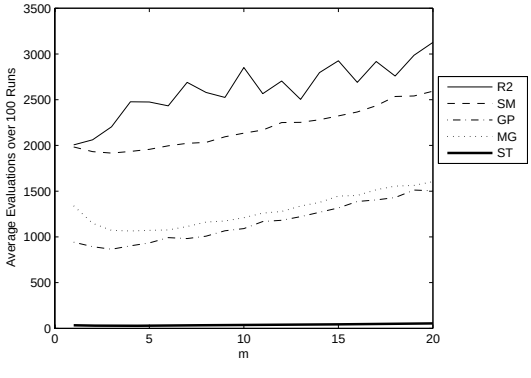


Fig. 4. Test Results of Parameter m

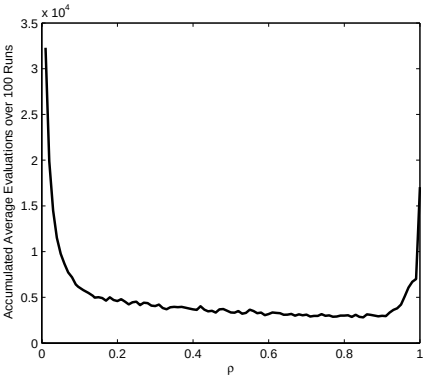


Fig. 5. Test Results of Parameter ρ

as ρ increases from 0.01 to about 0.9, but then increases as ρ increases. The proper setting of ρ would be 0.85 for all the subsequent tests.

4.2 Comparison with Other Ant System

Table 2 displays the test results of DACO comparing to other Ant Systems. Some information of the test results of other Ant Systems are unavailable, so the relative places are blank. The brackets indicate the tests are performed under a fixed evaluation number. For all the test functions, DACO outperforms other Ant Systems both in success rate and the average evaluation number. For all the functions tested, DACO finds all the known-optimal solutions. Overall, DACO greatly improves the ability of ACO in quickly finding the optimal solutions of function optimization problems with different characteristics.

Table 2. Comparison of DACO with Other Ant Systems

f	CACO		API		CIAC		ExtendedACO		DACO	
	% ok	evals	% ok	evals	% ok	evals	% ok	evals	% ok	evals
$R2$	100	6842		[10000]	100	11797		2905	100	1946.77
SM	100	22050		[10000]	100	50000		695	100	264.87
GP	100	5330			56	23391		364	100	229.53
MG	100	1688			20	11751			100	169.49
St		[6000]			94	28201			100	33.33

5 Conclusions

This paper presented DACO, a direct application of ACO to handle continuous optimization problems. In the proposed version of the system, pheromone trails represent the mean and deviation of the normal distribution for every variable,

and they are modified according to the previous search experience. Ants construct solutions by randomly selecting values for each variable biased by the normal distribution. A discussion of the speed of convergence, and some methods to escape from local optima are presented. Experimental results show that DACO outperforms other previous ACO related algorithms in success rate and the number of function evaluation. The results reported in the previous section demonstrate that DACO is capable of solving these various problems very rapidly and effectively.

The new version of the ACO algorithm improves the capabilities of the ACO to the function optimization field. Further research is under work for the principle parameter setting and quicker convergence speed.

References

1. G. Bilchev and I.C. Parmee. The ant colony metaphor for searching continuous design spaces. In T. Fogarty, editor, *AISB Workshop on Evolutionary Computing*, LNCS 933, pages 25–39, London, 1995. Springer-Verlag.
2. G. Bilchev and I.C. Parmee. Constrained optimization with an ant colony search model. In *2nd International Conference on Adaptive Computing in Engineering Design and Control*, pages 26–28, PEDC, University of Plymouth, 1996.
3. C. Blum and M. Dorigo. The hyper-cube framework for ant colony optimization. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 34(2):1161–1172, 2004.
4. E. Bonabeau, M. Dorigo, and G. Theraulaz. *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press, 1999.
5. M. Dorigo and L.M. Gambardella. Ant colony system: a cooperative learning approach to the traveling salesman problem. *IEEE Transactions on Evolutionary Computation*, 1(1):53–66, 1997.
6. M. Dorigo and T. Stutzle. *Ant Colony Optimization*. MIT Press, Boston, MA, 2003.
7. J. Dreo and P. Siarry. A new ant colony algorithm using the heterarchical concept aimed at optimization of multiminima continuous functions. In M. Dorigo, G. Caro, and M. Samples, editors, *Third International Workshop, ANTS 2002*, LNCS 2463, pages 216–221, Brussels, Belgium, 2002. Springer.
8. J. Dreo and P. Siarry. Continuous interacting ant colony algorithm based on dense heterarchy. *Future Generation Computer Systems*, 20(5):841–856, 2004.
9. L.M. Gambardella, E.D. Taillard, and G. Agazzi. Macs-vrptw: a multiple ant colony system for vehicle routing problems with time windows. In D. Corne, M. Dorgo, and F. Glover, editors, *New Ideas in Optimization*, pages 63–76. McGraw-Hill Ltd., London, U.K., 1999.
10. L.M. Gambardella, E.D. Taillard, and M. Dorigo. Ant colonies for the quadratic assignment problem. *Journal of the Operational Research Society*, 50(2):167–176, 1999.
11. M. Mathur, S.B. Karale, S. Priye, V.K. Jyaraman, and B.D. Kulkarni. Ant colony approach to continuous function optimization. *Ind. Eng. Chem. Res.*, 39(10):3814–3822, 2000.
12. N. Monmarche, G. Venturini, and M. Slimane. On how pachycondyla apicalis ants suggest a new search algorithm. *Future Generation Computer Systems*, 16(8):937–946, 2000.

13. K. Socha. Aco for continuous and mixed variable optimization. In M. Dorigo, editor, *ANTS 2004*, LNCS 3172, pages 25–36. Springer, 2004.
14. K.C. Tan, M.H. Lim, X. Yao, and L.P. Wang, editors. *Recent Advances in Simulated Evolution And Learning*. World Scientific, Singapore, 2004.
15. M. Wodrich and G. Bilchev. Cooperative distributed search: the ant's way. *Control & Cybernetics*, 3:413–446, 1997.
16. X. Yao. *Evolutionary Computation: Theory and Applications*. World Scientific, Singapore, 1999.

A Functions Tested in the Paper

Rosenbrock (R_n)	Variables: n	Minimum: $f(1,\cdots)=0$
$f(x)=\sum_{i=1}^{n-1}(100\cdot(x_i^2-x_{i+1})^2+(x_i-1)^2)$		
$x_i\in[-5,10]$		
Sphere Model (SM)	Variables: 6	Minimum: $f(0,\cdots)=0$
$f(x)=\sum_{i=1}^{6}x_i^2$		
$x_i\in[-5.12,5.12]$		
Goldstein & Price (GP)	Variables: 2	Minimum: $f(0,-1)=3$
$f(x)=(1+(x_1+x_2+1)^2\cdot(19-14x_1+3x_1^2-14x_2+6x_1x_2+3x_2^2))$		
$\cdot(30+(2x_1-3x_2)^2\cdot(18-32x_1+12x_1^2+48x_2-36x_1x_2+27x_2^2))$		
$x_i\in[-2,2]$		
Martin & Gaddy (MG)	Variables: 2	Minimum: $f(5,5)=0$
$f(x)=(x_1-x_2)^2+((x_1+x_2-10)/3)^2$		
$x_i\in[-20,20]$		
Step(S_t)	Variables: 5	Minimum: $f(x^*)=-30$
$f(x)=\sum_{i=1}^{5}[x_i]$ with $[x_i]$ greatest integer $\leq x$		
$x_i\in[-5.12,5.12]$		

Information Entropy and Interaction Optimization Model Based on Swarm Intelligence*

Xiaoxian He[1,2], Yunlong Zhu[1], Kunyuan Hu[1], and Ben Niu[1,2]

[1] Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang
[2] Graduate school of the Chinese Academy of Sciences, Beijing
{hexiaoxian, ylzhu}@sia.cn

Abstract. By introducing the information entropy *H(X)* and mutual information *I(X;Y)* of information theory into swarm intelligence, the Interaction Optimization Model (IOM) is proposed. In this model, the information interaction process of individuals is analyzed with *H(X)* and *I(X;Y)* aiming at solving optimization problems. We call this optimization approach as interaction optimization. In order to validate this model, we proposed a new algorithm for Traveling Salesman Problem (TSP), namely Route-Exchange Algorithm (REA), which is inspired by the information interaction of individuals in swarm intelligence. Some benchmarks are tested in the experiments. The results indicate that the algorithm can quickly converge to the optimal solution with quite low cost.

1 Introduction

As a new way of treating systems, swarm intelligence was noticed decades ago. In 1991, Marco Dorigo was inspired by the forage behaviors of ants, and proposed Ant Colony Optimizer (ACO) [1] for distributed optimization problems. After that, J. Kennedy and R. C. Eberhart were inspired by the forage behaviors of birds and fish, and proposed Particle Swarm Optimizer (PSO) [2] for continuous optimization problems. These methods have been applied to many problems and proved successful. Generally speaking, the Genetic Algorithm (GA) [3] is relevant to swarm intelligence, in that many individuals cooperate to solve problems by evolution. In addition, there are many analogous methods [4][5][6][11][12].

In the swarm systems, environment, individual and behavior rule are all sufficiently considered [7]. Based on them, however, systematical and general theory still cannot be established. In the past years, the information interaction of the individuals was ignored to some extend. At least there is not very elaborate analysis about it yet.

In this paper, we introduced the information entropy of Shannon [8] to depict the information of interaction among individuals, and proposed an interaction optimization model. Based on this model, a new algorithm is proposed to prove the model valid and effective. The rest of this paper is organized as follows. Section 2 gives a brief introduction to the swarm intelligence. Section 3 applies information entropy and mutual information to analyze the information interaction. Section 4 describes the

* This work is supported by the National Natural Science Foundation, China (No. 70431003) and the National Basic Research Program, China (2002CB312204).

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 136 – 145, 2006.

interaction optimization model. The new algorithm is proposed in section 5. Section 6 tests it on the benchmarks, and gives out the results. Finally, section 7 outlines the conclusions.

2 Swarm Intelligence

Swarm intelligence is defined as any attempt to design algorithm or distributed problem-solving devices inspired by the collective behaviors of the social insect colonies and other animal societies [9]. More generally speaking, swarm intelligence denotes that more complex intelligence emerges from the interactions of individuals who behave as a swarm. [1][9].

In a swarm intelligence model, environment, individual and behavior rule of individuals are three basic elements. All the individuals behave in given environments according to the rules. They are usually characterized by self-organization, and complex collective behavior emerges from the interaction of individuals. In social insects, individuals are self-autonomous. They obtain local information, and interact with their geographical neighbors. They also change the local environments or mark in the local environments to interact with the remote individuals indirectly, namely stigmergy.

When solving problems, however, there may be a lot of information each individual has. Some may be useful to his neighbors, while others not. Some given information may be useful to neighbor *a*, but it may be useless to neighbor *b*. As a result, the quality of information should be evaluated before interaction. This is the reason why information entropy must be applied.

3 Information Entropy and Information Interaction

Originally, the concept of entropy in thermodynamics referred to the amount of energy that is inaccessible for work. Shannon applied the concepts of information entropy and mutual information to the development of information theory of communication. High information entropy is equivalent to high levels of uncertainty or variability in a system. Information entropy has been broadly used to infer the uncertainty, disorganization, or variability in a physical system.

Basically, probabilities are the foundation of information entropy. Let $S = \{a_1, a_2, \ldots, a_M\}$ be a finite set and p be a proper probability density function on S. The amount of information needed to fully characterize all of the elements of this set consisting of M discrete elements is defined by

$$I(S_M) = \log_2 M \tag{1}$$

and is known as Hartley's formula. Shannon built on Hartley's formula, within the context of communication process, to develop his information criterion. His criterion, called entropy, is

$$H(X) = -\sum_{x \in S} p(x) \log p(x) \tag{2}$$

with *xlog(x)* tending to zero as *x* tends to zero. This information entropy measures the uncertainty or information content that is implied by *p*. *H(p)* reaches a maximum when $p_1 = p_2 = \cdots = p_M = 1/M$ (and is equal to Hartley's formula) and a minimum with $p(x_i) = 1, \exists x_i \in S$. It is emphasized here that *H(p)* is a function of the probability distribution. For example, if η is a random variable with possible distinct realizations $(x_1, x_2, \ldots, x_M)$ with probabilities $(p_1, p_2, \ldots, p_M)$, the entropy *H(p)* does not depend on the values $(x_1, x_2, \ldots, x_M)$ of η.

Considering two probability variants *X* and *Y*, if their joint probability density is $p(x, y)$, marginal probability is $p(x)$ and $p(y)$, respectively, and the conditional probability is $p(y \mid x)$, then the conditional entropy is defined as:

$$H(Y \mid X) = -\sum_{x,y} p(x,y) \log p(y \mid x) \tag{3}$$

Their mutual information $I(X;Y)$ is defined as:

$$I(X;Y) = \sum_{x \in S} \sum_{y \in S} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)} \tag{4}$$

If the information of Y is given, then it can be deduced as:

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

$$= \sum_{x,y} p(x,y) \log \frac{p(x \mid y)}{p(x)}$$

$$= -\sum_{x,y} p(x,y) \log p(x) + \sum_{x,y} p(x,y) \log p(x \mid y)$$

$$= -\sum_{x} p(x) \log p(x) - (-\sum_{x,y} p(x,y) \log p(x \mid y))$$

$$= H(X) - H(X \mid Y) \tag{5}$$

The result (5) indicates that the mutual information $I(X;Y)$ is the decrease of uncertainty for *X* when *Y* is given.

Since the information is finite in given environments, as for an individual, the smaller his information entropy is, the more information he has, and the more certain his behavior is. In other words, the information entropy is a measure for the quantity and quality of information. For a given individual, selecting which one to interact with can be determined by the mutual information besides time and

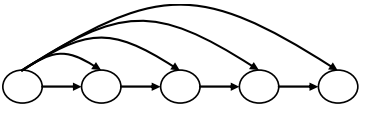
position. Interaction with those who can remarkably decrease uncertainty can accelerate to solving the problems.

4 Interaction Optimization Model

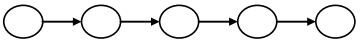
Assuming each individual has a finite set of actions $A=\{a_1, a_2, \ldots, a_N\}$ available to it, and $S=\{s_1, s_2, \ldots, s_N\}$ is a finite state space. Formally, an action a maps a_k state s_i into another state s_j. Whether to take the action is determined by the behavior rules, the past states, and the heuristic information at discrete-time stochastically. Therefore, an action can be represented by a set of transition matrices with elements p_{ij}^{kt}. This is the probability of entering state s_j at time *t+1* given that at time *t* the *k*th individual was in state s_i and an action was executed, or mathematically

$$p_{ij}^{kt} = \Pr ob(S^{t+1} = s_j \mid S^t = s_i, S^{t-1} = s_{i-1}, \ldots, S^0 = s_0) \tag{6}$$

The selection of the next state may depend not only on the current state, but also the past state the individual has experienced, namely the system is not necessarily Markovian. In this paper we treat on the general stochastic process. Fig.1 shows the two kinds of stochastic process.



(a) A general stochastic process



(b) A Markov chain

Fig. 1. Two kinds of stochastic process

Definition 1: If the *k*th individual is in state s_i at time *t*, then the information entropy is:

$$H_i^{kt} = -\sum_{j \in allowed_i^{kt}} p_{ij}^{kt} \log p_{ij}^{kt} \tag{7}$$

where $allowed_i^{kt}$ denotes the set of all candidate states when the individual at the state s_i.

Definition 2: Considering two state sequences *A* and *B*, if there exists the same state sequence as *A* in *B*, then we call $A \subset B$.

Definition 3: Considering the *k*th individual and *l*th individual, the state sequences they experienced are S_k and S_l, respectively. If their initial state is the same one, and $S_k \subset S_l$, then the information of the *k*th individual is redundant to the *l*th individual.

Deduction 1: If the information of the *k*th individual is redundant to the *l*th individual, and vice versa, then the state sequences the two have experienced are identical.

Definition 4: Let $p(y^{lt})$ be the distribution of the *l*th individual at time *t*, if his information has influence on the *k*th individual's action, then $p(x^k \mid y^{lt})$ denotes the distribution of the *k*th individual after the information of the *l*th is given, otherwise

$$p(x^k \mid y^{lt}) = p(x^k) \tag{8}$$

Definition 5

$$p(x^{kt}, y^{lt}) = p(y^{lt})p(x^{kt} \mid y^{lt}) \tag{9}$$

It should be noticed that

$$p(x^{kt}, y^{lt}) \neq p(y^{lt}, x^{kt}) \tag{10}$$

Now we can give the definition of mutual information.

Definition 6

$$I(X^{kt};Y^{lt}) = \sum_{x^{kt},y^{lt}} p(x^{kt}, y^{lt}) \log \frac{p(x^{kt}, y^{lt})}{p(x^{kt})p(y^{lt})}$$

$$= \sum_{x^{kt},y^{lt}} p(x^{kt}, y^{lt}) \log \frac{p(x^{kt} \mid y^{lt})}{p(x^{kt})}$$

$$= H(X^{kt}) - H(X^{kt} \mid Y^{lt}) \tag{11}$$

$I(X^{kt};Y^{lt})$ measures the decrease of uncertainty for the *k*th individual at time *t* after the information of the *l*th individual is given.

Assuming the heuristic information is always helpful to improve the individuals' fitness, and the rules are the same for each individual. In the process of search, the individual should evaluate its information itself firstly. As for the *l*th individual, the smaller the value of $H(Y^{lt})$ is, the better the quality of its information is, and the more reliable the information is for other individuals with whom the *l*th individual would interact. Then, the value of $I(X^{kt};Y^{lt})$ determines whether the *k*th individual should interact with the *l*th. The larger the value of $I(X^{kt};Y^{lt})$ is, the more influence the *k*th individual has on the *l*th. In this way, the individuals can be more and more intelligent to find the solution. The process is illustrated in Fig. 2.

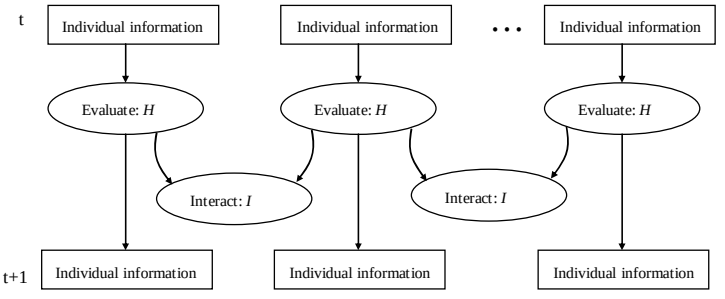


Fig. 2. The interaction optimization model

5 The Route-Exchange Algorithm

The route-exchange algorithm is designed for solving Traveling Salesman Problem (TSP). The problem is to find the shortest route for a traveling salesman to visit all the cities once and only once, and return to the starting city, which is also known as a Hamiltonian cycle. Because the exchange of route information is the most important idea in the algorithm, it is called route-exchange algorithm. The problem can be described as follows.

Given a set of cities $C = \{c_1, c_2, \cdots, c_n\}$, for each pair (c_i, c_j), $i \neq j$, let $d(c_i, c_j)$ be the distance between city c_i and c_j. Solving the TSP entails finding a permutation π^* of the cities $(c_{\pi^*(1)}, \cdots, c_{\pi^*(n)})$, such that

$$\sum_{i=1}^{n} d(c_{\pi^*(i)}, c_{\pi^*(i+1)}) \leq \sum_{i=1}^{n} d(c_{\pi(i)}, c_{\pi(i+1)}), \forall \pi \neq \pi^*, (n+1) \equiv 1 \qquad (12)$$

In the symmetric TSP $d(c_i, c_j) = d(c_j, c_i), \forall i, j$, while in the asymmetric TSP this condition is not satisfied. In this work we consider the symmetric TSP.

The route-exchange algorithm is a heuristic approach inspired by the behaviors of people, in which every individual searches the state space independently and simultaneously. As far as people are concerned, if a number of them are asked to fulfill the same task independently, they would learn from each other to improve their own performances when encountering. This mechanism is applied to the route-exchange algorithm.

Given a symmetric TSP of n nodes, m individuals are required to find a closed tour independently from different starting nodes with the same velocity v, where $m \leq n$. They could encounter the fellows somewhere in the process. Each of them has two storages: one is for the permutation of cities $(c_{\pi(1)}, \cdots, c_{\pi(n)})$ visited by itself in the current repetition, another is for the best permutation $(c_{\pi^*(1)}, \cdots, c_{\pi^*(q)})$

visited by itself or others ever encountered in the current repetition, where $q \le n$. For the convenience of depiction, we denote the former with *route-storage-A, RS_A* for short, and the latter with *RS_B.* When two individuals encounter each other, they would compare the quality of the two routes visited in the current repetition with each other, put the better one into *RS_B*, and continue to tour. In the next repetition, the individual would prefer to *RS_B* of previous repetition and follow it first, then previous *RS_A*.

$$p_{ij}^{k} = \begin{cases} \dfrac{(\eta_{ij})^{\alpha} \cdot \rho_B}{\sum_{r \in allowed_k} (\eta_{ir})^{\alpha}}, l_{ij} \in RS\_B_k \\ \dfrac{(\eta_{ij})^{\alpha} \cdot \rho_A}{\sum_{r \in allowed_k} (\eta_{ir})^{\alpha}}, (l_{ij} \in RS\_A_k) \wedge (l_{ij} \notin RS\_B_k) \\ \dfrac{(\eta_{ij})^{\alpha} \cdot \rho_C}{\sum_{r \in allowed_k} (\eta_{ir})^{\alpha}}, else \end{cases} \tag{13}$$

As for the *k*th individual, if he is at node *i* sometime, then the probability that he selects the node *j* as his next node is defined as (13), where $\eta_{ij} = 1/d_{ij}$, l_{ij} denotes the edge (i, j), $allowed_k$ denotes the set of all candidate nodes when the *k*th individual is at node *i*, ρ_A, ρ_B and ρ_C are selection parameters, $\rho_C < \rho_A < \rho_B$, and α is a control parameter.

Considering the fact of the real world and the feasibility of computation, we take time that individuals reach the same node as interaction criteria instead of *I(X;Y)*. Given that the *k*th individual reaches the node *i* at time T_k, and the *l*th individual reaches the same node at time T_l. If the following inequality is satisfied, then the two individuals are judged encountered.

$$\left|T_k - T_l\right| \le \frac{\overline{d_{ij}}}{\beta v}, j \in Neighbor_i \tag{14}$$

where $Neighbor_i$ is the set of nodes connected directly with node i, and β is a control parameter.

In order to evaluate the quality of the route $(c_{\pi(1)}, \cdots, c_{\pi(q)})$, we have *G* that

$$G = H^{kt} \cdot \frac{\sum_{i=1}^{q-1} d_{i,i+1}}{q} \tag{15}$$

where H^{kt} is the information entropy of the *k*th individual at time t. The smaller the value of *G* is, the better the quality of the route is.

A positive feedback mechanism is designed to avoid vibrations. After each repetition, the individual having the best results in the current repetition saves his experience, and take the same tour in the next repetition. In this way, the best individual can keep his capability of influence on others, and accelerate others to improve their fitness. The pseudo-code of route-exchange algorithm is illustrated in Fig. 3.

```
Step 1   Initialize, put all individuals on different nodes randomly, NC=1
Step 2   Take the nodes as the starting nodes, tour the cities by (13) with velocity v
Step 3   Repeat Step 2 for each individual; Goto Step 4
Step 4   For each node, calculate the time T that every individual reaches it, determine
         whether two individuals encounter with each other according to (14)
           if yes, compare their routes with G in (15) and put the better routes into
         RS_B
Step 5   Find the individual who has the best results in the current repetition, put his
         RS_A into RS_B
Step 6   NC=NC_max?
         if yes, stop the program and return the results
         else, RS_B is empty?
               if yes, put the individuals on the first node of their RS_A
               else put the individuals on the last node of RS_B
                              end
                  end
         Goto Step 7
Step 7   NC=NC+1, Goto Step 2
```

Fig. 3. The route-exchange algorithm

6 Experiments and Results

The route-exchange algorithm was tested on some TSP instances defined in the TSPLIB [10]. The experiments were conducted on several Windows XP Pentium IV 1.60 GHz PCs with 256 Mbytes of memory. The results of the instances are summarized in Table 1. The first column stands for the names of the test instances, the second for exact optimal tour length for each problem given in the TSPLIB. The third column stands for the best results we have obtained, the fourth to the sixth for the results in each run time, the seventh for the average results, and the eighth for the relative error (Err) respectively, where the relative error is calculated as

$$Err = \frac{Ave - Opt}{Opt} \times 100\% \tag{16}$$

The ninth column denotes the repetition times required, and the last denotes the number of individuals employed in the program.

Table 1 indicates that the proposed route-exchange algorithm can be used to solve symmetric TSP effectively. It can be seen that among 8 test problems, the maximum relative error is 8.22% (of the test problem Eil51) and the average relative error of all is easily calculated to be 4.71%. While the number of nodes increases, the repetition times required doesn't increase remarkably. In other words, the results can converge quickly to good solutions, and the complexity of the computation can be kept quite low.

Table 1. Summary of the experiment results

Problem	Opt	Best Result	Run1	Run2	Run3	Ave.	Err(%)	NC_{max}	Ind.
Burma14	33.23	31.88	33.23	31.88	33.21	32.77	-1.38	10	10
Ulys-ses22	75.67	75.67	76.06	76.63	75.67	76.12	0.59	30	20
Bayg29	9074	9399	9399	9519	9433	9450	4.14	30	20
Bays29	2020	2097	2134	2097	2107	2113	4.60	30	28
Swiss42	1273	1325	1325	1374	1357	1352	6.21	30	30
Att48	33524	35073	35073	36573	37073	36240	8.10	30	40
Eil51	426	451	451	473	459	461	8.22	30	40
Berlin52	7542	7870	8182	7870	8208	8087	7.23	30	40

7 Conclusions

In the biological world, information interaction phenomenon exists widely in most species from insects to mammals. It is necessary for them to fulfill tasks cooperatively. In this paper we introduced the information entropy and mutual information into the artificial intelligence system, which can be applied to analyze this phenomenon mathematically. On the basis, the interaction optimization model is proposed as a general model for swarm intelligence. This is also an endeavor to descript the swarm intelligence theoretically. Then a new algorithm is proposed to validate it. As can be seen in the experiments and results, this algorithm can converge quickly to a good solution with quite low computing complexity. The stagnancy in local optimum, which is generally existent as a defect in the analogous other algorithms, can be effectively avoided. Also, the positive feedback mechanism we designed plays a very important role in this algorithm, which helps to attract other individuals to the global optimum step by step.

References

1. Bonabeau, E., Dorigo, M., Theraulaz, G.: Swarm Intelligence-from Natural to Artificial System. Oxford University Press, New York (1999)
2. Kennedy, J., Eberhart, R.C.: Particle Swarm Optimization. In: Proceedings of IEEE International Conference on Neural Networks, Piscataway, NJ (1995) 1942-1948

3. Grefenstette, J., Gopal, R., Rosimaita, B., Van Gucht, D.: Genetic algorithms for the traveling salesman problem. In: Proceedings of the International Conference on Genetics Algorithms and their Applications (1985) 160-168
4. Yao, X.: Evolutionary Computation: Theory and Applications. World Scientific, Singapore (1999)
5. Tan, K.C., Lim, M.H., Yao, X., Wang L.P. (Eds.): Recent Advances in Simulated Evolution and Learning. World Scientific, Singapore (2004)
6. Liu, J., Zhong, W.C., Liu, F., Jiao, L.C.: Organizational coevolutionary classification algorithm for radar target recognition. Journal of Infrared and Millimeter Waves 23(3) (2004) 208-212
7. Han, J., Cai, Q.S.: Emergent Intelligence in AER Model. Chinese Journal of Pattern Recognition and Artificial Intelligence 15(2) (2002) 134-142
8. Shannon, C. E.: A mathematical theory of communication. Bell System Technology Journal 27 (1948) 397-423
9. Kennedy, J., Eberhart, R.C., Shi, Y.: Swarm Intelligence. Morgan Kaufmann Publishers, San Francisco (2001)
10. http://www.iwr.uni-heidelberg.de/groups/comopt/software/TSPLIB95
11. Niu, B., Zhu, Y.L., He, X.X.: Multi-population Cooperative Particle Swarm Optimization. Lecture Notes in Artificial Intelligence 3630 (2005) 874-883
12. Niu, B., Zhu, Y.L., He, X.X.: A Multi-Population Cooperative Particle Swarm Optimizer for Neural Network Training. Lecture Notes in Computer Sciences 3971 (2006) 570-576

PSO with Improved Strategy and Topology for Job Shop Scheduling

Kun Tu[1,2], Zhifeng Hao[1,3], and Ming Chen[3]

[1] College of Computer Science and Engineering, South China University of Technology, Guangzhou 510640, P.R. China
[2] College of Mathematical Science, South China University of Technology, Guangzhou 510640, P.R. China
[3] National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, P.R. China
captorxp@163.com

Abstract. Particle swarm optimization (PSO) has proven to be a promising heuristic algorithm for solving combinatorial optimization problems. However, N-P hard problems such as Job Shop Scheduling (JSSP) are difficult for most heuristic algorithms to solve. In this paper, two effective strategies are proposed to enhance the searching ability of the PSO. An alternate topology is introduced to gather better information from the neighborhood of an individual. Benchmarks of JSSP are used to test the approaches. The experiment results indicate that the improved Particle Swarm has a good performance with a faster searching speed in the search space of JSSP.

1 Introduction

The job shop scheduling problem (JSSP) is one of the most difficult N-P hard problems in the sense of its high computational complexity [1]. The JSSP is described as follow: There are n jobs with m operations scheduled on m machines in a given order. Each operation is specified to a fixed machine and has a fixed processing time. The constraints of the problem are [7]:

(1) Each job must pass through each machine once and only once.
(2) Each job should be processed through the machine in a particular order.
(3) Each operation must be executed uninterrupted on a given machine.
(4) Each machine can only handle at most one operation at a time.

It is impractical to use some traditional exact algorithms to solve this problem even in a small size because of the time-consuming computational procedure. The constraints of the problem are so strong, making the valid search space of the problem too complicated.

Recently, many heuristic algorithms have been developed to obtain the solution. The performances are quite good because those algorithms can obtain the near optimal or even the best solution with useful techniques. Among them, genetic algorithm (GA) is one of the most promising techniques to solve N-P hard problem. It is usually combined with local search methods to get a near optimal result [1].

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 146 – 155, 2006.

Particle Swarm Optimization (PSO) is a newly developed algorithm by Eberhart and Kennedy [8]. Later, many useful techniques and mechanism have been added to PSO to improve its computational efficiency. Several effective topology modals for the population of PSO such as 'Star', 'Triangle' and 'Ring' etc are proposed [4, 5]. Later, a hybrid PSO was proposed to solve the JSSP and gained a good result [17, 18].

However, PSO does not always gain an efficient performance in the minimization of some discrete multimodal functions because of the different searching methods. This paper studies the searching strategy of PSO, tries to explain the reason why some topologies are better than others and then presents two improved strategies and an alternate topology.

2 Job Shop Scheduling Model for the Algorithm

2.1 The Representation

There are several kinds of representations for JSSP such as priority-rule-based representation, random key representation and operation-based representation [15]. This paper uses the job-order-based representation because it is easy for the PSO to compute. For example, a solution to a 3-job and 2-machine problem may be expressed in a $_{3\times 2}$ integer array as [1 3 3 2 1 2], where the value of the component in the array represents the job number. Thus, the array above means that the first operation of job 1 should be processed first, followed by the first and the second operation of job 3. The whole procedure should end with the last operation of job 2.

The position found by an individual may be invalid, so for an n-job and m-machine problem, the array should be repaired to satisfy the constraints:

(1) The value of the component should not be less than 1 or larger than n;

(2) There are m components having the same integer value of i, where $i = 1, 2, \cdots, n$.

2.2 The Left Shift for the Representation

When the operation-based representation is adopted, there exist two kinds of solution spaces: semi-active solution space and active solution space. In a semi-active solution space, some solutions can be decoded to semi-active schedules in which a certain operation can be put ahead without modifying the processing time of other operations. In an active solution space, the schedule corresponding to an active solution cannot be modified by putting an operation ahead.

The left shift operator changes a semi-active schedule into an active schedule. The maximum job completion time can be improved when the maximum finishing time of the jobs is shortened.

This paper implements the technique of reconstructing a semi-active solution into an active one in [14] to enhance the performance of the algorithm.

3 Particle Swarm Optimization

The idea of PSO originates from the social behavior of swarms of birds or fish schools. It has been reported that each bird or fish in the swarm in a school moves according to the information given by others in the neighborhood. Similar to the GA, there is a population called a swarm in PSO, and every individual in the swarm is called a particle which represents a solution to the problem. In an iteration, each particle moves in the search space from one position to another according the information from its neighborhood. The distance for which the particle moves in one iteration is called velocity and the position of the particle is evaluated by an object function. Each particle records the best position it found in the previous iteration as a local best. A global best position is chosen from the local bests of all the particles.

Converting a solution to an n-dimensional vector, the search space S is accordingly an n-dimensional space, where $S \in R^n$. The position $X_i = (x_{i1}, x_{i2}, \cdots, x_{in})$ of the i th particle is a point in S, the velocity $V_i = (v_{i1}, v_{i2}, \cdots, v_{in})$ of the i th particle is also a vector with n dimension, The local best position of the i th particle denotes as $P_i = (p_{i1}, p_{i2}, \cdots, p_{in})$ .The best of the local best positions is chosen as the global best position $P_g = (p_{g1}, p_{g2}, \cdots, p_{gn}) = \max\{P_i\}$.

The object function is used to evaluate the quality of the position X found by a particle, denoted as $f(X)$. In the t th iteration, each particle updates it position with the following formula:

$$V_{t+1} = \chi(V_t + \varphi(P_m - X_t)) \tag{1}$$

$$X_{t+1} = X_t + V_t \tag{2}$$

where $P_m = (\varphi_1 \cdot P_i + \varphi_2 \cdot P_g)/(\varphi_1 + \varphi_2)$ and $\varphi = \varphi_1 + \varphi_2$, φ_1 and φ_2 are the values of the acceleration coefficient limits, χ is the constriction coefficient determined by φ_1 and φ_2.

Kennedy et al changed the swarm topology and neighborhood information by modifying the form of P_m [6]:

$$P_m = \frac{\sum_{k \in N} W(k)\varphi_k P_k}{\sum_{k \in N} W(k)\varphi_k} \tag{3}$$

where φ_k is an random value between 0 and $\varphi_{\max} / |N|$, N is the set of neighbors of the particle and P_k is the best position found by individual k. $W(k)$, the weight function of the information given by individual k, is usually related to a certain aspect of the particle such as the fitness of the best position or the distance from particle k to the current individual.

4 Improved PSO

4.1 Analysis of the Topology

It is pointed out that the increase of neighborhood size of a particle may deteriorate the performance of the swarm [6]. 'UAll' and 'All' topologies obtained worse results than other topologies such as Square and Ring versions.

In this paper, we try to explain the reason why the truly fully informed version such as All topology has a worse performance. In the experiments, we found that not all the information gathered from the neighborhood contributes to the individual's movement. In some cases of the 'All' topology, the velocity V_{t+1} even leads the particle to the opposite direction of the global best or the local best. Although this phenomenon is important for a particle to search for a new position, the fitness of the new position is always lower than expected. In formula (3), the larger $|N|$ is, the more information an individual gets from the neighborhood. For those particles whose best position ever found are above the average, much of the information from the worse positions in the neighborhood is not worth using because it leads the individual to a worse direction. The more the useless information is, the more "confused" an individual will be. From Kennedy's experiments, we can deduce that appropriate amount of information sources is better than the truly informed version. Thus, the refining of the information from the neighborhood is needed for an individual to improve its direction.

There are two tasks left for a particle before it moves to a new position with a better strategy:

(1) Selecting useful information from the neighborhood to determine a proper direction to a better position;
(2) Evaluating the velocity of last iteration and make adjustment according to the past experiences as well as the information in the neighborhood.

4.2 Information Selecting

The original PSO uses information from the best position in the population and the individual's best. Chunming et al uses the information of worse positions to keep away from them [9]. These two strategies suggest simply that a better position potentially provides positive instruction. Thus, adding good particles to the neighborhood may produce an efficient topology. On the other hand, a worse position is more likely to mislead the individual to an area with low fitness value. This suggests a direct way to refine the information.

Here, an alternate form of P_m is proposed to implement the information selecting:

$$P_m = \frac{\sum_{i \in N^-} W(i)\varphi_i P_i + \sum_{j \in M} W(j)\varphi_j P_j}{\sum_{k \in N^- \cup M} W(k)\varphi_k} \tag{4}$$

where N^- is the set of particles better than the individual in the neighborhood and M is the set of the three best position in the population.

4.3 Velocity Evaluating and Adjustment

Formula (1) shows how a particle moves in an iteration: it keeps part of its previous velocity with inertia and adjusts its direction to a new position according to the information from the neighborhood. However, the individual moves without "questioning" whether it will go to an area with low fitness value in the search space. As has mentioned in the analysis of the topology, one of the tasks of an individual is to check the quality of the previous velocity so as to make an adjustment.

One applicable method to evaluate the velocity is to see whether the individual has found a better position than before. If the fitness of the positions improves iteration after iteration, the particle probably moves towards a direction to an optimum. On the other hand, if the fitness decreases continuously, the particle may go in a wrong direction and just search in an area with low fitness.

Two methods can avoid the individual to go towards the potentially inappropriate area: (1) "jump over" the area and (2) "step back" a little. The weight of V_t in formula (1) can be changed adaptively to implement the adjustment:

$$V_{t+1} = \chi(\alpha V_t + \varphi(P_m - X_t)) \tag{5}$$

$$\alpha = \begin{cases} a & fitness(i)_{t-j} < fitness(i)_{t-j-1} \\ 1 & otherwise \end{cases} \tag{6}$$

where $a > 1/\chi$ in "jump over" version and $a = -1$ in "step back" version, $0 \le j \le j_{\max}$, $fitness(i)_{t-j}$ is the fitness of the position found by individual i in the $t-j$ iteration. Formula (6) indicates that an individual should avoid the inertia of previous velocity if the fitness continuously decreases for j iterations.

5 Experiments and Results

Two versions of the PSO were tested and compared with the GA. The PSO with the "jump over" strategy is denoted as JOPSO and the "step back" version is denoted as SBPSO. One experiment is to test the contribution of the strategies to the searching speed; another one is carried out to test the efficiency of the improved PSO.

The parameters of the PSO are set as follow according to [6]: the population is 100, $j_{\max}$ is 3 and a is set to be 3 in JOPSO, χ is 0.7298 and $\varphi_{\max}$ is 2.05. $W(i) = 1/fitness(i)$.

The same operators of GA as in [14] are implemented. The parameters are as follow: population size is 100; crossover probability is 0.8; mutation probability is 0.2. All the algorithms will stop after 3000 iterations or if the global best stops improving for 500 iterations.

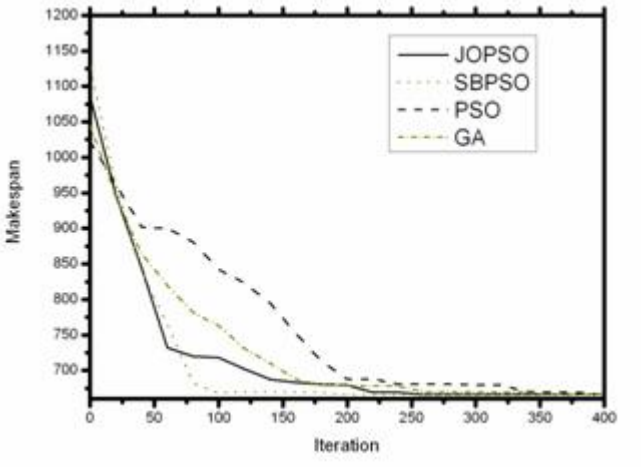


Fig. 1. Performance of the algorithms on La01

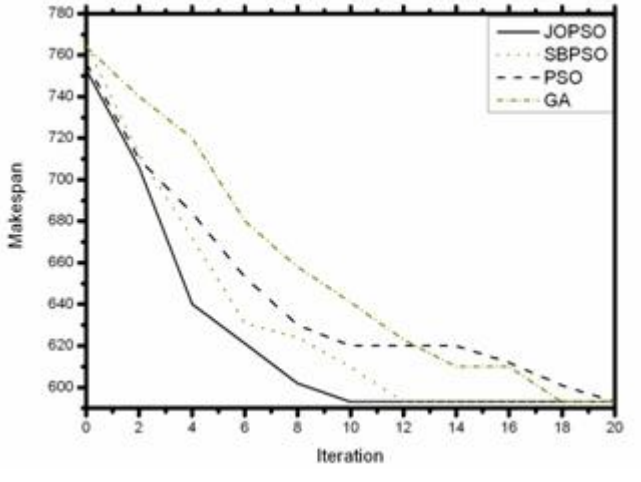


Fig. 2. Performance of the algorithms on La05

In the evaluation of the strategies, we measure the searching speed of the algorithms by iterations. Although the computation time shows directly the speed at which an algorithm can find an optimal of a problem, it is not applicable when the

algorithms are implemented on different machines. Meanwhile, it is hard to illustrate clearly the effect of a strategy to a heuristic algorithm in a step searching for a better solution. In the tests, the global best of the population of the algorithms is recorded in each iteration. Benchmarks of JSSP for OR-library from [16], namely, la01, la05 and abz5 are used to test the searching speed of the PSOs and the GAs.

In Figure 1, JOPSP had a faster searching speed than the other three in the first 75 iterations. However, SBPSO had a better performance than JOPSO after about 75 iterations and found the global optimum of 666 before 200 iterations. The JOPSO converged slower than the SBPSO but faster than the GA and the traditional PSO, reaching the optimum after 200 iterations. The result can be explained as follow: in the solution space of la01 problem, the "step back" strategy is less effective when a better solution can be easily found at the beginning of the search; but its contribution to the performance of PSO is steadier than "jump over" strategy when better solutions are hard to find.

Figure 2 and figure 3 suggest that JOPSO also has a good searching ability and performed the best. It must be pointed out that the la05 problem is relatively easy for most algorithms to find the optimal solution. Thus, the "jump over" strategy can boot up the speed of the algorithm.

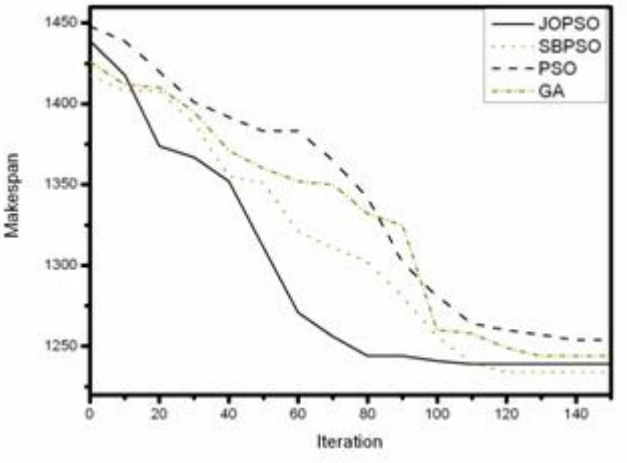


Fig. 3. Performance of the algorithms on Abz5

Next, more benchmarks with larger size are used to test the efficiency of the PSOs and the GAs, namely, mt10, mt20, abz7, abz8, la21 and la22. Each benchmark is tested for 40 times on every algorithm, the best and the worst results are recorded and then the average of the 40 results is calculated. The proposed PSOs are compared to the fGA and GA-asol reported in [14].

Table 1 illustrates the experimental results of fGA, the problem sizes and the optimum of each benchmark are also listed in it. Table 2, Table 3 and Table 4 show the results of GA-asol, JOPSO and SBPSO.

Table 1. Experimental results of fGA

Benchmarks	n	m	optimum	Best	Worst	Average
Mt10	10	10	930	937	982	957.33
Mt20	20	5	1165	1184	1234	1207.47
Abz5	10	10	1234	1234	1253	1246.15
Abz7	20	15	656	716	744	731.85
Abz8	20	15	(645,669)	742	769	754.66
La21	15	10	1046	1082	1130	1104.61
La22	15	10	927	948	986	970.85

Table 2. Experimental results of GA-asol

Benchmarks	Best	Worst	Average
Mt10	930	967	949.25
Mt20	1173	1199	1182.38
Abz5	1234	1250	1241.58
Abz7	700	726	715.57
Abz8	720	746	733.61
La21	1074	1109	1092.10
La22	940	970	953.52

Table 3. Experimental results of JOPSO

Benchmarks	Best	Worst	Average
Mt10	955	965	958.83
Mt20	1165	1184	1171.53
Abz5	1234	1242	1240.21
Abz7	666	716	683.74
Abz8	681	703	697.49
La21	1047	1082	1064.04
La22	927	940	936.88

Table 4. Experimental results of SBPSO

Benchmarks	Best	Worst	Average
Mt10	942	955	951.71
Mt20	1165	1179	1172.46
Abz5	1234	1250	1239.8
Abz7	675	703	680.27
Abz8	681	710	686.49
La21	1054	1073	1062.19
La22	927	935	934.22

From the four tables, the improved PSOs gain better results than the GAs in the test. In half of tested problems, the best solutions found by the PSOs are the optima, while the GAs only found two. The worst results by the PSOs are also impressive: some solutions are the same as the best ever found by the GAs(see Mt20, Abz7 and Abz8).

The “best” column in Table 3 and Table 4 illustrate that JOPSO found more good solution than SBPSO, which means JOPSO has a better searching ability than SBSO. However, results of the SBPSO from the “worst” and “average” columns are better. This suggests that the “step back” strategy is steadier than the “jump over” version.

It must be noted that the improved PSOs do not always gain better results than the GAs. In the mt10 benchmark, neither of the PSOs found a good solution in the 40 times of computation. The PSOs are probably misled to a local optimum in the solution space.

6 Conclusion

In this paper, an improved topology and two alternative searching strategies were proposed specifically for PSO to solve discrete N-P hard problem such as JSSP. The experimental results show that the searching speed of PSO was improved by the topology and the new strategies. The strategies enhance the searching ability of the algorithm by leading the particles to a potentially better direction. The “jump over” strategy performs better than the “step back” version in some benchmarks where better solution is easy to find in the solution space, while in other benchmarks the “step-back” shows its efficiency in the steady converging speed.

An exception of the good performance of the improved method is showed in the testing of mt10 benchmark. This phenomenon suggests that a different method should be used for a different problem. That is, there probably exists no such a powerful algorithm that can solve all the problems better than others. An effective method for some problems may perform badly in other problems.

The two strategies are actually adaptive techniques for the parameter adjustment of the PSO. The heuristic algorithm relies a lot on its parameters to gain a good result. Thus, the parameters should be further investigated according to a specific problem.

Acknowledgement

This research is supported by open research fund of National Mobile Communications Research Laboratory (No.200605)).

References

[1] Van Laarhoven P.J.M., E.H.L.Aarts, J.K.Lenstra, "Job Shop Scheduling by Simulated Annealing Oper. Res., 40(1992), 113-125

[2] Aihua Yin, Wenqi Huang, ” A stochastic strategy for solving job shop scheduling problem”, *Proceeding of the First International Conference of Machine Learning and Cybernetics*, 2002.

[3] Hongfang Zhang, Xiaoping Li, Pin Zhou, "A Job Shop Oriented Virus Genetic Algorithm", *Proceeding of the 5th World Congress on Intelligent Control and Automation*, June 2004.
[4] James Kennedy, Rui Mendes, "*Population Structure and Particle Swarm Performance*", IEEE 2002
[5] James Kennedy, "Small Worlds and Mega-Minds: Effects of Neighborhood Topology on Particle Swarm Performance", IEEE 1999.
[6] James Kennedy, Jose Neves, "The Fully Informed Particle Swarm: Simpler, Maybe Better", *IEEE Transactions on Evolutionary Computation*, Vol.8, NO.3, June 2004
[7] French, S.: Sequencing and Scheduling: An introduction to the Mathematics of the Job-Shop. *John Wiley & Sons, Inc.*, New York (1982).
[8] R. C. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proc. 6th Symp. MicroMachine and Human Science*, Nagoya, Japan, 1995, pp. 39–43.
[9] Chunming Yang; Simon, D.; A new particle swarm optimization technique; *Systems Engineering, 2005. ICSEng 2005*. 18th International Conference on16-18 Aug. 2005 Page(s):164 - 169
[10] Dorndorf and E. Pesch, Evolution based learning in a job-shop environment, *Computers and Operations Research,* Vol. 22, pp.25-40, 1995.
[11] Bean, Genetic algorithms and Random Keys for sequencing and optimization. *ORSA J. Computing,* Vol. 6, pp. 154-160, 1994.
[12] Bierwirth, A generalized permutation approach to job-shop scheduling with genetic algorithms. *OR SPEKTRUM,* Vol. 17, NO. 2-3, pp.87-92, 1995.
[13] Zhang Hong-Fang; Li Xiao-Ping; Zhou Pin; A job shop oriented virus genetic algorithm; *Intelligent Control and Automation, 2004. WCICA 2004. Fifth World Congress* Volume 3, 15-19 June 2004 Page(s):2132 - 2136 Vol.3
[14] Masato Watanabe, Kenichi Ida, Mitsuo Gen; Active solution space and search on job-shop scheduling problem; *Electrical Engineering in Japan, Volume 154, Issue 4, Date: March 2006, Pages: 61-67*
[15] Wang Ling. Shop Scheduling with Genetic Algorithm. Tsinghua University Press,2003
[16] Mattfeld DC, Vaessens RJM. OR-Library. http://mscmga.ms.ic.ac.uk/jeb/orlib/ jobshopinfo .html
[17] Wei-jun Xia, Zhi-ming Wu. A hybrid particle swarm optimization approach for the job-shop scheduling problem. *The International Journal of Advanced Manufacturing Technology*. 2005. Pp: 1433-3015
[18] Bo Liu, Ling Wang, Yi-Hui Jin. An effective hybrid particle swarm optimization for no-wait flow shop scheduling. *The International Journal of Advanced Manufacturing Technology*. Jan 2006. pp: 1433-3015.

Virus-Evolutionary Particle Swarm Optimization Algorithm

Fang Gao[1], Hongwei Liu[1], Qiang Zhao[2], and Gang Cui[1]

[1] School of Computer Science and Technology, Harbin Institute of Technology, 150001 Harbin, China
gaofang@hit.edu.cn, {liuhongwei, cg}@ftcl.hit.edu.cn
[2] School of Traffic, Northeast Forestry University, 150040 Harbin, China
qyangzhao@163.com

Abstract. This paper presents an improved discrete particle swarm optimization algorithm based on virus theory of evolution. Virus-evolutionary discrete particle swarm optimization algorithm is proposed to simulate co-evolution of a particle swarm of candidate solutions and a virus swarm of substring representing schemata. In the co-evolutionary process, the virus propagates partial genetic information in the particle swarm by virus infection operators which enhances the horizontal search ability of particle swarm optimization algorithm. An example of partner selection in virtual enterprise is used to verify the proposed algorithm. Test results show that this algorithm outperforms the discrete PSO algorithm put forward by Kennedy and Eberhart.

1 Introduction

The particle swarm optimization (PSO) algorithm, proposed by Kennedy and Eberhart [1,2], is a new optimization technique originating from artificial life and evolutionary computation [3,4]. The algorithm completes the optimization through following the personal best solution of each particle and the global best value of the whole swarm. PSO can be implemented with ease and few parameters need to be tuned. It has been successfully applied in many areas and proved to be a very effective approach in the global optimization for complicated problems [5,6]. However, as a newly developed optimal method, the PSO algorithm is still in its developmental infancy and further studies are necessary. For example, the original PSO had difficulties in controlling the balance between exploration and exploitation because it tends to favor the intensification search around the 'better' solutions previously found [7]. In such a context, the PSO appears to be lacking global search ability. Also, its computational efficiency for a refined local search to pinpoint the exact optimal solution is not satisfactory [8,9].

Many researchers have proposed various kinds of approaches to improve the PSO algorithm [10,11,12]. To enhance the performance of PSO, we first propose a virus-evolutionary particle swarm optimization algorithms (VEPSO). It is composed of a particle swarm and a virus swarm. The process of the VEPSO is similar to co-evolution based on both a horizontal propagation between particle and virus particles and a vertical inheritance of genetic information from ancestors to offspring. In the

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 156 – 165, 2006.

co-evolve process, the virus particles propagate partial genetic information in the particle swarm by virus infection operators. Numerical examples presented by us show that the model has better global convergence performance.

This paper is organized as follows. Section 2 presents a new PSO algorithm based on virus theory of evolution. The virus infection operators are defined and incorporated into the PSO. Section 3 presents an application of VEPSO in partner selection of virtual enterprise problem. In section 4, the effectiveness of the proposed algorithm is shown through some numerical simulations. Further, the comparison between the VEPSO and the binary PSO proposed by Kennedy and Eberhart is discussed through simulation results.

2 Virus-Evolutionary Particle Swarm Optimization Algorithm

2.1 Overview of PSO Algorithm

Like genetic algorithm, the PSO algorithm first randomly initializes a swarm of particles. Each particle is represented as $X_i = (x_{i,1}, x_{i,2}, ..., x_{i,n})$, $i = 1,2,...,N$, where N is the swarm size. Thus, each particle is randomly placed in the n-dimensional space as a candidate solution. Each particle adjusts its trajectory toward its own previous best position and the previous best position attained by any particle of the swarm, namely $pbest_i$ and $gbest$. In each iteration, the swarm is updated by the following equations:

$$v_{i,j}^{k+1} = v_{i,j}^{k} + c_1 r_1 (pbest_{i,j}^{k} - x_{i,j}^{k}) + c_2 r_2 (gbest_j^{k} - x_{i,j}^{k}), \quad (1)$$

$$x_{i,j}^{k+1} = x_{i,j}^{k} + v_{i,j}^{k+1}, \quad (2)$$

where k is the current iteration number, $v_{i,j}$ is the updated velocity on the j th dimension of the i th particle, c_1 and c_2 are acceleration constants, r_1 and r_2 are real numbers drawn from two uniform random sequences of $U(0, 1)$.

The above continuous particle swarm algorithm has been used to solve many optimization problems. In the discrete binary version [1], a particle moves in a state space restricted to zero and one on each dimension, where each $v_{i,j}$ represents the probability of bit $x_{i,j}$ taking the value 1. Thus, the step for updating $v_{i,j}$ remains unchanged as shown in Eq. (1), except that $pbest_{i,j}$ and $gbest_j$ are integers in {0, 1} in binary case. The resulted changes in position are defined as

$$s(v_{i,j}^{k+1}) = 1/(1+\exp(-v_{i,j}^{k+1})), \quad (3)$$

$$\text{if } (r < s(v_{i,j}^{k+1}) \text{ then } x_{i,j}^{k+1} = 1, \text{ else } \quad x_{i,j}^{k+1} = 0 \quad (4)$$

where r is random number drawn from uniform sequence of $U(0,1)$.

2.2 Virus-Evolutionary Particle Swarm Optimization Algorithm

Virus theory of evolution is based on the view that virus transduction is a key mechanism for transporting segments of DNA across species [13,14,15,16,17,18]. A virus-evolutionary PSO algorithm simulates co-evolution based on the virus theory of evolution. VEPSO has two swarms: a particle swarm and a virus swarm (Fig.1). The particle swarm and the virus swarm are defined as a set of candidate solutions and a substring set of the particle swarm, respectively. We assume the virus infection enables the horizontal propagation of genetic information in the intra-generational particle swarm, while the updating of particles' position and velocity enable the vertical inheritance of genetic information from particles' current generation to their next generation. VEPSO has two virus infection operators as follows [17]:

(a) Reverse transcription operator (infection): A virus overwrites its substring on the string of particles for generating new particles (Fig.2.a).
(b) Incorporation operator (transduction): A virus takes out a substring from the string of a particle for generating new virus particles (Fig.2.b).

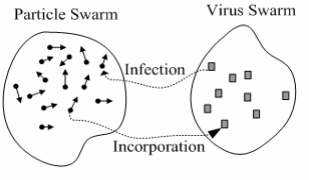


Fig. 1. VEPSO algorithm

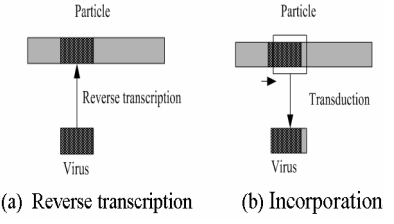


Fig. 2. Virus infection operators

First, a virus performs the reverse transcription to a particle randomly selected out of the particle swarm. Each virus has $fitvirus_i$, as a strength about the virus infection. We assume that $fitparticle_j$ and $fitparticle'_j$ are the fitness value of a particle j before and after the infection, respectively. The $fitvirus_{i,j}$ denotes difference between $fitparticle_j$ and $fitparticle'_j$, which is equal to the improvement obtained by infecting to the particle as

$$fitvirus_{i,j} = fitparticle'_j - fitparticle_j \tag{5}$$

$$fitvirus_i = \sum_{j \in S} fitvirus_{i,j} \tag{6}$$

where i is the virus number and S is a set of particles which are infected by the virus i.

After infections, each virus evolves by incorporation according to $fitvirus_i$. If $fitvirus_i$ is larger than zero, the virus performs incorporation by taking out partially new substring from one of the infected particle. Otherwise, a virus shortens the genotype by removing some substrings. Actually, the virus length shortens/extends according to an incorporation probability (P_i). P_i denotes the probability to take out a substring from a particle. In addition, each virus has $infrate_i$ for the virus infection as [14]:

$$infrate_{i,k+1} = \begin{cases} (1+\alpha) \cdot infrate_{i,k} & if \quad fitvirus > 0 \\ (1-\alpha) \cdot infrate_{i,k} & otherwise \end{cases} \tag{7}$$

where α (>0) is coefficient. When $fitvirus_i$ is high, $infrate_i$ also becomes high. The increase of $infrate_i$ accelerates the increase of effective schemata by virus infection. Furthermore, each virus has a life force as follows:

$$life_{i,k+1} = r \times life_{i,k} + fitvirus_i \tag{8}$$

where k and r mean the generation and the life reduction rate, respectively. If r takes a negative value, the virus particle takes out a new substring with the incorporation operator from the randomly selected particle. The VEPSO's procedure is as follows:

```
begin
t=0;
Initialize particle and virus swarm respectively;
t=1;
repeat
   Update particle（position，velocity）;
   Virus infect;
   Evaluate fitness of particle;
   Evaluate fitness of virus;
   t=t+1;
until (Termination condition is true)
end
```

In the above procedure, initialization randomly generates an initial particle swarm, and then a virus is generated as a substring of a particle encoding. The virus randomly takes out each substring from the particle according to probability P_i. The worst particle with least fitness is eliminated as the selection scheme. New particles generated by virus infections, as new individuals. After virus infections, if the fitness of one particle is improved comparing with its current particle, then the current position is replaced with its infected position. Consequently, the successfully infected particles survive into the next generation. This indicates the virus infection operator can be regarded as a local hill-climbing operator in the long run.

3 An Example: Partner Selection of Virtual Enterprise

3.1 Problem Definition

Suppose that a virtual enterprise consists of n business processes (core competencies): $E_i, i = 1,2,...,I$. For the i th business process, there are a_i potential candidates for the partners. Let m_j^i, $j = 1,2,...,a_i$, be the j th potential candidate for the i th business process. The partner selection problem can be described as follows: take one candidate from m_j^i for each business process E_i to form a virtual enterprise, for

example $m_3^1 m_1^2 m_1^3 m_2^4 m_4^5$, and the resultant combination of selected partners must satisfy an optimization objective.

Since the running cost, reaction time and running risk are the key points in the operation of virtual enterprise, we take them as the optimization objectives for partner selection[19,20]:

(1) Running cost: namely C, consists of the internal cost of each candidate and the link cost between any two candidates:

$$\min C = \min\left\{\sum_{i=1}^{I}\sum_{j=1}^{a_i}(\beta_j^i C_j^i) + \sum_{i=1}^{I}\sum_{j=1}^{a_i}\sum_{i'=1,i'\neq i}^{I}\sum_{j'=1}^{a_i}(\beta_j^i \beta_{j'}^{i'} C_{j,j'}^{i,i'})\right\} \tag{9}$$

where $\beta_j^i = \begin{cases} 1 & m_j^i \text{ is selected} \\ 0 & m_j^i \text{ is not selected} \end{cases}$, C_j^i is the internal cost of the candidate m_j^i. $C_{i,i'}^{j,j'}$ is the link cost between two candidates of m_j^i and $m_{j'}^{i'}$.

(2) Reaction time: namely T, consists of the internal reaction time of each candidate and the link time between any two candidates:

$$\min T = \min\left\{\sum_{i=1}^{I}\sum_{j=1}^{a_i}(\beta_j^i T_j^i) + \sum_{i=1}^{I}\sum_{j=1}^{a_i}\sum_{i'=1,i'\neq i}^{I}\sum_{j'=1}^{a_i}(\beta_j^i \beta_{j'}^{i'} T_{j,j'}^{i,i'})\right\} \tag{10}$$

where T_j^i is the internal reaction time of the candidate m_j^i. $T_{i,i'}^{j,j'}$ is the link time between two candidates of m_j^i and $m_{j'}^{i'}$.

(3) Running Risk: namely R, is defines as follows:

$$\min R = \min\left\{\sum_{i=1}^{I}\max_j(\beta_j^i R_j^i)\right\} \tag{11}$$

where R_j^i is the running risk generated by selecting m_j^i.

Here we take the weighted sum of Eq. (9), (10) and (11) as the objective function:

$$F(X) = \frac{\omega_c}{E_c}\left(\sum_{i=1}^{I}\sum_{j=1}^{a_i}(\beta_j^i C_j^i) + \sum_{i=1}^{I}\sum_{j=1}^{a_i}\sum_{i'=1,i'\neq i}^{I}\sum_{j'=1}^{a_i}(\beta_j^i \beta_{j'}^{i'} C_{j,j'}^{i,i'})\right)$$

$$+\frac{\omega_t}{E_t}\left(\sum_{i=1}^{I}\sum_{j=1}^{a_i}(\beta_j^i T_j^i) + \sum_{i=1}^{I}\sum_{j=1}^{a_i}\sum_{i'=1,i'\neq i}^{I}\sum_{j'=1}^{a_i}(\beta_j^i \beta_{j'}^{i'} T_{j,j'}^{i,i'})\right) + \frac{\omega_r}{E_r}\left(\sum_{i=1}^{I}\max_j(\beta_j^i R_j^i)\right) \tag{12}$$

where E_c, E_t and E_r are the least desired values of C, T and R respectively, and ω_c, ω_t and ω_r are the weight coefficients of C, T and R respectively.

3.2 Particle Representation and Fitness Function Selection

Suppose n is the total number of all the business process types. Since each particle of the swarm is one of the candidate solutions for the problem, we represent each particle by a binary vector as follows:

$$X = \{b^i\} = \{b^1, b^2, ..., b^n\}, \text{ subject to } \sum_{j=1}^{a_i} b_j^i \geq 1 \tag{13}$$

where $b^i = \{b_1^i, b_2^i, ..., b_{a_i}^i\}$, as a bit in X, b_j^i represents the j th candidate for the i th process, which is restricted to take integers in {0, 1} according to

$$b_j^i = \begin{cases} 1 & \text{if } m_j^i \text{ is selected} \\ 0 & \text{if } m_j^i \text{ is not selected} \end{cases} \tag{14}$$

Based on the optimization objective function in Eq. (12), the fitness function $f(x)$ of each particle is set as $f(x) = 1/F(x)$.

3.3 Virus Infection

A virus individual can transmit a substring between particle individuals. A substring of a virus individual consists of three characters {0,1,*} and is the same length as that of the particle individual. The character '*' denotes 'don't care' mark. A virus individual does not perform the reverse transcription in the same position where there is a '*'. In this case, the length of the virus is constant, but the order of the virus is variable. Figure 3 shows an example of the reverse transcription.

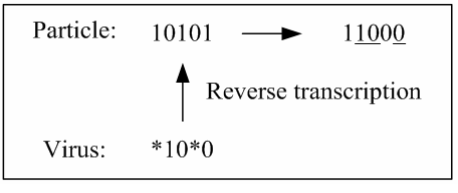
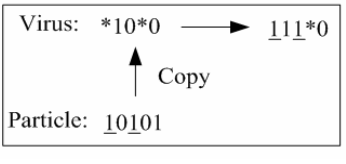
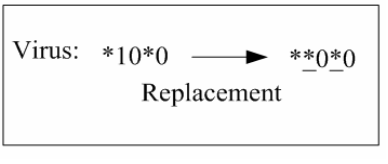


Fig. 3. Reverse transcription operator

The reverse transcription overwrites the substring of the virus individual on a randomly selected particle individual. The incorporation has two types of operations (Fig. 4). The first one is to copy substrings from a particle. The other is to replace some substrings with character '*'. If the virus improves the fitness values of particle individuals, the copy operator is performed with probability P_i, otherwise, the replacement operator is performed.



(a) Copy operator



(b) Replacement operator

Fig. 4. Incorporation operators

3.4 Optimization Procedure

The procedure of the proposed VEPSO algorithm for solving the optimization problem of partner selection is described as follows:

Step 1. Initialize particle swarm: Generate N particles with random positions: $X_1, X_2, ..., X_N$, according to Eq. (1). Generate velocities $v_{i,j}$, $i = 1,2,...,N$, and $j = 1,2,...,n$, where $v_{i,j} \sim U(0,1)$.

Step 2. Initialize virus swarm: Generate $0.1 \times N$ viruses randomly.

Step 3. Evaluate each particle according to $f(x) = 1/F(x)$ based on Eq. (12).

Step 4. Evaluate each virus according to Eq. (5) and (6).

Step 5. Individual and global best positions updating: If $f(pbest_i) < f(X_i)$, then $pbest_i = X_i$, and search for the maximum value $f_{\max}$ among $f(pbest_i)$, If $f(gbest) < f_{\max}$, $gbest = X_{\max}$, $X_{\max}$ is the best particle associated with $f_{\max}$.

Step 6. Velocity updating: update the i th particle velocity using the Eq. (1) restricted by maximum and minimum threshold $v_{\max}$ and $v_{\min}$.

Step 7. Position updating: update the i th particle position using Eq. (3) and (4).

Step 8. Virus infection using Eq. (7) and constraint checking. If constraint is not satisfied, reset to step 7 and repeat step 8 until it is satisfied.

Step 9. Virus life force computation using Eq. (8).

Step 10. Repeat step 3 to 9 until a given maximum number of iterations is achieved.

4 Numerical Simulations

In this section we use a detailed example of partner selection to illustrate the effectiveness of the VEPSO algorithm. A supply chain for some product consists of four business processes: design (D), purchase (P), manufacture (M) and sale (S) . Candidates for each business process are presented in Table 1. Our goal is to select one candidate from each business process to minimize the objective functions in Eq. (12). The internal running cost, reaction time and running risk of each candidate are presented in Table 2. The link time and cost between candidates are as shown respectively in Table 3. In Table 3, elements on the left lower region of the diagonal line represent the link time, and elements on the right-upper region represent the link cost between two candidates.

Table 1. Business processes E_i and their candidate enterprises m_j^i

Business Processes	Design		Purchase					Manufacture				Sale		
Candidates	D_1	D_2	P_1	P_2	P_3	P_4	P_5	M_1	M_2	M_3	M_4	S_1	S_2	S_3

Table 2. Internal cost C_j^i, reaction time T_j^i and running risk R_j^i of each candidate enterprise

	D_1	D_2	P_1	P_2	P_3	P_4	P_5	M_1	M_2	M_3	M_4	S_1	S_2	S_3
$T_j^{(i)}$ /month	8.8	5.3	7.6	4.5	8.9	10.5	11.8	4.4	9.7	2.4	7.8	9.1	8.3	5.9
$C_j^{(i)}$ /10^4 RMB	85.3	74.6	55.5	82.6	91.2	63.2	51.6	99.4	96.5	86.4	92.9	89.7	90.2	84.0
R_j^i	0.3	0.5	0.5	0.2	0.2	0.4	0.4	0.3	0.4	0.2	0.5	0.3	0.3	0.2

Table 3. Link cost $C_{i,i'}^{j,j'}$ (10^4 RMB) and link time $T_{i,i'}^{j,j'}$ (month) between any two candidates

	D_1	D_2	P_1	P_2	P_3	P_4	P_5	M_1	M_2	M_3	M_4	S_1	S_2	S_3
D_1	0	5.3	1.7	2.5	5.8	5.0	5.7	3.2	0.7	3.9	2.3	3.2	3.8	3.7
D_2	5.9	0	2.1	1.6	5.9	4.2	4.1	2.3	4.8	3.6	5.9	2.1	1.2	4.1
P_1	1.9	5.1	0	4.1	2.3	1.4	4.1	5.0	1.6	5.9	3.5	5.9	3.0	3.3
P_2	5.7	1.6	3.7	0	1.1	0.6	1.3	2.0	1.0	4.7	1.1	3.7	1.6	5.5
P_3	5.7	3.3	3.1	2.6	0	3.6	0.8	0.9	4.5	4.3	1.4	4.8	0.6	2.1
P_4	4.1	3.6	0.0	3.8	3.5	0	0.9	5.5	4.5	0.7	4.9	5.1	5.7	4.5
P_5	3.0	2.1	2.4	2.0	4.4	4.9	0	2.6	5.0	1.5	1.1	3.2	3.6	3.5
M_1	0.4	5.2	4.8	5.8	5.1	2.6	4.7	0	5.5	4.5	3.1	0.6	4.8	5.8
M_2	5.9	1.4	1.6	1.5	0.2	0.2	5.9	0.5	0	3.2	3.3	5.7	0.9	0.7
M_3	2.0	1.6	5.2	0.5	1.0	1.7	4.7	3.2	0.8	0	0.9	5.2	0.6	3.5
M_4	3.8	3.2	5.0	4.1	0.0	4.6	2.2	0.8	4.2	2.2	0	1.3	5.6	4.3
S_1	5.6	3.0	2.2	2.8	2.6	1.6	5.3	2.6	4.2	3.9	3.8	0	4.9	5.0
S_2	4.3	5.6	5.8	2.1	2.6	4.8	5.5	1.8	4.2	6.0	3.5	1.8	0	3.6
S_3	3.1	0.5	5.7	5.0	3.3	2.0	0.1	4.7	3.3	3.3	1.5	0.7	5.7	0

Parameters of the problem are set as: $E_c = 150$, $E_t = 60$, $E_r = 0.2$, $\omega_c = 0.25$, $\omega_t = 0.45$ and $\omega_r = 0.3$ .Parameters of the algorithm are set as: $N = 100$, $c_1 = 2$, $c_2 = 2$, $v_{\max} = 4$ and $v_{\min} = -4$. Main virus parameters are set as: virus swarm size equals to 10, *infrate*=0.02, α =0.9 and P_i=0.1. Fig. 5-(a) shows a typical running of the VEPSO algorithm and it shows that the objective function $F(x)$ quickly converges to the optimal solution as the iteration generation increases

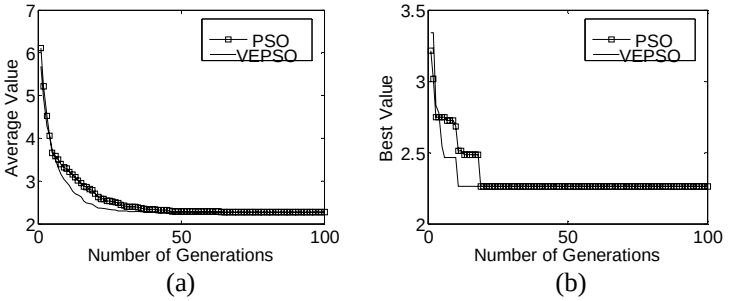


Fig. 5. Iteration process of the algorithm

The optimum represented by *gbest* equals to {10010000010001} with the objective functions $F(gbest) = 2.2625$, which means that the best partner combination is {D_1, P_2, M_3, S_3}. Fig. 5-(b) shows that the average value of the whole swarm descends and approaches quickly to the best value in Fig. 5-(a).

Simulation with the discrete PSO proposed by Kennedy and Eberhart (here called PSO) is also performed as a comparison with our VEPSO algorithm. We set the PSO swarm size and other parameters to be same with the swarm size of the VEPSO. The results are also presented in Fig. 5. It shows that the VEPSO algorithm has faster converging speed and better average value than the discrete PSO proposed by Kennedy and Eberhart. To present detail comparison further, the above two algorithms are run fifty times respectively, and the corresponding statistical results is shown in Table 4. From this table, it can be seen that VEPSO algorithm outperforms the PSO algorithm.

Table 4. Statistical contrast results of algorithms

Algorithms	Mean of number of generations	Standard deviation of number of generations
PSO	10.62	4.28
VEPSO	7.78	3.79

5 Conclusions

This paper presents a virus-evolutionary particle swarm optimization algorithm, which has two features of horizontal propagation and vertical inheritance of genetic information. VEPSO enhances the global search ability and the refined local search ability of available PSOs. Furthermore, an example of partner selection of virtual enterprise is presented to show the VEPSO algorithm's effectiveness and efficiency.

Acknowledgement

This work is supported by the National Science Foundation of China under grant No. 60503015.

References

1. Kennedy, J., Eberhart, R.: Particle Swarm Optimization. In: Proceedings of the IEEE International Conference on Neural Networks, Vol. 4. Perth. Australia. (1995) 1942–1948
2. Eberhart, R., Kennedy, J.: New Optimizer Using Particle Swarm Theory, in Proc. 6th Int. Symp. Micro Machine Human Science. (1995) 39–43
3. Yao, X.: Evolutionary Computation: Theory and Applications. World Scientific, Singapore (1999)
4. Tan, K.C., Lim, M.H., Yao, X., Wang L.P. (Eds.): Recent Advances in Simulated Evolution and Learning. World Scientific, Singapore (2004)

5. Zhao, Q., Yan, S.Z.: Collision-Free Path Planning for Mobile Robots Using Chaotic Particle Swarm Optimization. Lecture Notes in Computer Science.3612(2005)632–635
6. Li, Y.M., Chen, X.: Mobile Robot Navigation Using Particle Swarm Optimization and Adaptive NN. Lecture Notes in Computer Science.3612(2005)628–631
7. Silva, A., Neves, A., Costa ,E.: An Empirical Comparison of Particle Swarm and Predator Prey Optimization. Lecture Notes in Computer Science. 2464(2002) 103–110
8. Schutte, J. F., Groenword, A. A.: A Study of Global Optimization Using Particle Swarms. J. Global Optimiz., 31(2005) 93–108
9. Ho,S. L., Yang, S.Y., Ni, G.Z., Wong, H. C.:A Particle Swarm Optimization Method with Enhanced Global Search Ability for Design Optimizations of Electromagnetic Devices. IEEE Transations on Magnetics. 42(2006) 1107–1110.
10. Lu ,Z., Hou, Z. : Particle Swarm Optimization with Adaptive Mutation. Acta Electronca Sinica. 3(2004) 417–420
11. Jiang, C., Etorre, B. : A Self-adaptive Chaotic Particle Swarm Algorithm for Short Term Hy-droelectric System Scheduling in Deregulated Environment. Energy Conversion and Man- agement. 46(2005) 2689–2696
12. Chatterjee, A., Siarry, P.: Nonlinear Inertia Weight Variation for Dynamic Adaptation in Particle Swarm Optimization. Computers & Operations Research. 33(2006) 859–871
13. Kubotan ,N., Koji, S. et: Role of Virus Infection in Virus-evolutionary Genetic Algorithm. Proceedings of the IEEE Conference on Evolutionary Computation. (1996) 182–187
14. Kubotan, N., Fukuda, T. et.: Virus-evolutionary Genetic Algorithm for a Self-organizing Manufacturing System. Computers ind. Engng. 30(1996) 1015–1026
15. Kubotan, N., Fukuda ,T. et.: Trajectory Planning of Cellar Manipulator System Using Virus-Evolutionary Genetic Algorithm. Robotics and Autonomous System. 19(1996) 85–94
16. Kubotan, N., Fukuda, T. et.: Evolutionary Transition of Virus-evolutionary Genetic Algorithm. Proceedings of the IEEE Conference on Evolutionary Computation. (1997) 291–296
17. Kubotan, N., Arakawa ,T. et.: Trajectory Generation for Redundant Manipulator Using Virus Evolutionary Genetic Algorithm. Proceedings of the IEEE Conference on Robotics and Automation. (1997) 205–210.
18. Kubotan, N., Fukuda, T.: Schema Representation in Virus-Evolutionary Genetic Algorithm for Knapsack Problem. Evolutionary Computation Proceedings,1998, IEEE World Congress on Computational Intelligence – The 1998 IEEE International Conference on. Anchorage, USA: IEEE. (1998) 834–839
19. Feng, W. D., Chen, J., Zhao, C. J.: Partners Selection Process and Optimization Model for Virtual Corporations Based on Genetic Algorithms. Journal of Tsinghua University(Science and Technology). 40(2000) 120–124
20. Qu, X. L., Sun, L. F.: Implementation of Genetic Algorithm to the Optimal Configuration of Manufacture Resources. Journal of Huaqiao University. 26(2005) 93–96

Intelligent Particle Swarm Optimization Algorithm and Its Application in Optimal Designing of LPG Devices for Optical Communications Fields

Yumin Liu[1,2] and Zhongyuan Yu[1,2]

[1] School of Science, Beijing University of Posts and Telecommunications, 100876, Haidian District Beijing, China
[2] Key Laboratory of Optical Communication and Lightwave Technologies, Ministry of Education 100876, Haidian District Beijing, China
liuyuminhqy@263.net, yuzhongyuan30@hotmail.com

Abstract. An innovation long period grating (LPG) synthesis method based on intelligent particle swarm optimization (PSO) algorithm is demonstrated to be very effective for designing optimized LPG filters. A flatten 3dB loss spectrum LPG is designed as an example to show the effectiveness of the PSO algorithm. This 3dB LPG is a key component of multi-channel filter in optical communications and optical sensors. The results showed that the intelligent PSO algorithm is very powerful and can be used for complex optimization problem.

1 Introduction

Fiber gratings have been evolved into one of the key optical components and found a multitude of applications in optical communications and optical sensor fields [1,2]. LPG is one of the most popular transmission-type grating devices, which couples the fundamental guided mode to one or more the phase-matched cladding modes. The LPG filters have proved to be very useful in band rejection filter [3], high sensitive temperature and strain sensors [4], and EDFA gain flattening [5]. Generally specking, the uniform LPG has little applications because of the limited transmission loss spectrum. In recent years, lots of synthesis methods are developed to design the index modulation profile that can produce the specific spectrum applications. These synthesis methods can be roughly divided into two groups: the fist group is inverse scattering based algorithm including the layer-peeling method, which can be used for both the LPG and fiber Bragg grating (FBG) designing [6,7]. For a reasonable spectrum, these methods can uniquely determine the index modulation. However, it is generally not easy to determine weather the spectrum is reasonable or not, and the couple strength generally should not too strong or there will be large divergence. The other kind of synthesis methods is the stochastic optimization approaches. Because the FBG or LPG filters synthesis is generally a complex optimization problem, so the used optimization algorithms are not common sense's optimization algorithms. The widely used optimization algorithms for complex problem are evolution algorithms that originate from the bioscience. The advantages of these methods are that we can obtain an easily implemented index modulation profile by imposing additional constraints on the solution.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 166 – 175, 2006.

The evolutionary programming (EP) and genetic algorithm (GA) are important branches of the evolution algorithms, which are probabilistic search algorithms gleaned from the organic evolution process [8,9]. EP is relative simple compared with the GA) and only use the mutation process of continuous variables and does not use the coding and crossover process. However, for both these methods, the controllable parameters are problem sensitive and not easy selective properly. Recently, PSO has been proposed as a simple and alternative tool for solving optimization problems. PSO has been shown to be effective in optimization for multidimensional discontinuous problems. PSO has also been successfully applied in electromagnetic design problems. In this article, we have successfully utilized this intelligent optimization algorithm to design a LPG with flatten 3dB loss spectrum within the bandwidth about 100nm. In another example, we designed a single-stage long period grating EDFA gain flattening filter for the whole C-band.

We organized the papers as follows: first we gave the principle of the PSO algorithm and show how to improve PSO algorithm by some modification to the original version; second we gave a simple revive about the LPG filters, and focused on how to design of LPG by PSO algorithm; third, we demonstrated an design example to prove the effectiveness of the improved PSO and the design results are also shown with some useful discussions; in the final, conclude remarks are given.

2 The Principle of PSO

The PSO algorithm, like GA, is population-based optimisation tool and can be applied to virtually any problem that can be expressed in terms of an objective function for which extremum must be found [10]. It emulates some aspects of social behaviour of a flock of birds and a school of fish. The PSO algorithm is iterative and involves initializing a number of vectors (particles) randomly within the search space. The collective of the particles is known as the swarm. Each particle presents a potential solution to the problem of the target function. At initial stage, each particle is also randomly initializing a vector called particle speed. During each time step (generation), the objective function is evaluated to establish the fitness of each particle using the particle position itself as the input parameter. Then the particle will fly through the search space being attracted to both their personal best position as well as the best position found by the swarm so far.

We make the position of particle i expressed as $X_i = (x_{i1,}, x_{i2} \ldots, x_{id})$ and the best position of that particle so far expressed $p_i = (p_{i1}, p_{i2} \ldots, p_{id})$. The best position of the whole swarm can be expressed as $p_g = (p_{g1}, p_{g2}, \ldots, p_{gd})$. Then the particle position update can be expressed as $x_{id} = x_{id} + v_{id}$ where the v_{id} denote the speed of the d dimension component of particle i and expressed as:

$$v_{id} = wv_{id} + \phi_1 rand()(p_{id} - x_{id}) + \phi_2 rand()(p_{gd} - x_{gd}) \quad (1)$$

where w is the inertia weight determining how much of the particle's previous speed is preserved, ϕ_1, ϕ_2 are two acceleration constants present the cognition part and the social part respectively, $rand()$ is uniform random sequences from $\{0,1\}$.

It should be noted that, each particle has the memory ability and could remember the best position it has been gone through. What's more, there is information flow among the particles, by which they could communicate with the global particles or with the special particles based on some topological neighbourhood structure. Each particle updates its speed and position based on its memory and communion information of the swarm. So we called the PSO algorithm as intelligent PSO algorithm.

The absolute value of the current speed determines the searching depth and expandability in solve spacing. The direction of the speed determines the path by which the particle could quickly or slowly go to the best solution. So the speed update is very important in the PSO algorithm. Fro the speed update, there are two basic versions: one employ the weight factor W, A large weight facilities the global exploration, while a small exploration facilities the local exploration; the other one employ the K factor, which has proved to be better than the first one in success probability and searching speed [8]. Thus, in later sections, the position vectors and speed vectors in our paper are mainly based on the constriction factor factor K technique, which describes in detailed as follows:

$$v_{id} = \begin{cases} K\left(v_{id} + \phi_1 rand()\left(p_{id} - x_{id}\right)\right. \\ \quad \left. + \phi_2 rand()\left(p_{gd} - x_{id}\right)\right), \ x_{\min} < x_{id} < x_{\max} \\ v_{id} = 0, \ \ otherwise \end{cases} \tag{2}$$

$$x_{id} = \begin{cases} x_{id} + v_{id}, & x_{\min} < x_{id} < x_{\max} \\ x_{\max,} & x_{id} \geq x_{\max} \\ x_{\min,} & x_{id} \leq x_{\min} \end{cases} \tag{3}$$

where K is computed as: $K = 2\left|2 - \varphi - \sqrt{\varphi^2 - 4\varphi}\right|^{-1}$, $\varphi = \phi_1 + \phi_2$, $\varphi > 4$. ϕ_1, ϕ_2 are constants represents the social learning and self recognition components and set 1.1 and 3.0 respectively in our algorithm. Just as can be seen in (2), we used the absorber boundary conditions when the one components of the particle vector overcome the boundary. In other PSO literatures, boundary selection problem have been studied and show that reflection and trap boundary generally can produce better performance.

The iterative process will continue using the formula (2) until the extremum has been found or the number of iteration reached the maximum value. The algorithm in pseudo-code follows:[11]

Intialize population
Do
For $i = 1$ to population swarm size
If $f\left(\vec{x}_i\right) < f\left(\vec{p}_i\right)$ then $\vec{p}_i = \vec{x}_i$
$\vec{p}_g = \min\left(\vec{p}_{neighors}\right)$

for $d = 1$ to Dimension

$$v_{i,d} = wv_{i,d} + \varphi_1 rand()\left(p_{i,d} - x_{i,d}\right) + \varphi_2 rand()\left(p_{g,d} - x_{i,d}\right)$$

$$v_{i,d} = sign\left(v_{i,d}\right)\min\left(abs(v_{i,d}), v_{\max}\right)$$

$$x_{i,d} = x_{i,d} + v_{i,d}$$

Next d

Next i

Until termination criterion is met

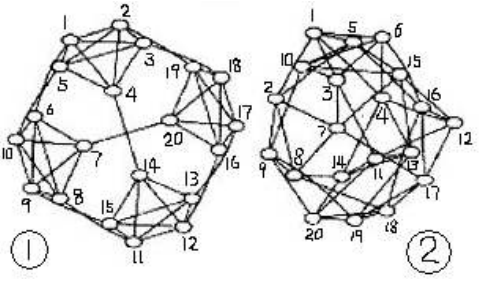


Fig. 1. Two possible particles swarm local topological structures used for LPSO

The ability of avoiding being trapped into the sub optimal is an important factor to evaluate the performance of an algorithm. Fro the PSO algorithm, it provides the selective topological structures to solve this problem. The two most common used structures are known as a global neighbor and a local neighbor. In the global neighborhood structure, the trajectory of each particle's search is influenced by the best position found so far and the other particles of the entire swarm. The local neighborhood structure allows each individual particle to be influenced by only some small number of adjacent members. A kind of lore has evolved regarding these sociometric structures. It has been thought that the global structure converges rapidly on problem solutions but has a weakness of being rapidly on problem solutions but has a weakness of being trapped into local optima, while the local structure can flow around local optima and avoids being trapped into local optima. In this paper we use two possible structures just as in reference [7,9] Generally, there are two kinds of topological neighborhood structures, global neighborhood structure, corresponding to the global version of PSO (GPSO), and local neighborhood structure, corresponding to the local version of the PSO (LPSO). For the global neighborhood structure the whole swarm is considered as the neighborhood, while for the local neighborhood structure some smaller number of adjacent members in sub-swarm is taken as the neighborhood. In the GPSO, each particle's search is influenced by the best position found by any member of the entire population. In contrast, in LPSO, the search is influenced only by parts of the adjacent members. It is widely believed that GPSO converges quickly to an optimum but has the weakness of being trapped in local optima occasionally, while the LPSO is able to "flow around" local optima, because

the sup-swarm explore different regions in hyperspace. To improve the search ability, for the simple applications, we can use the GPSO algorithm and get the optimized solution in short searching time. For the complex applications, we used the LPSO algorithm to improve the probability of finding the best global solution. The two possible topologies for LPSO are shown in Fig.1 [12].

3 Analysis and Optimization of LPG

The coupled mode equations describing the coupling in long period fibre gratins are given by [13]

$$\begin{aligned} dA/dz &= -ikB - i\delta A \\ dB/dz &= -ikA + i\delta B \end{aligned} \tag{4}$$

where A B are respectively the amplitude of the guide mode and cladding mode, k is the couple strength, $\delta = \pi\Delta_{neff}/\lambda - \pi/\Lambda$ represents the detuning and $\Delta_{neff} = n_{cor} - n_{cla}$, where n_{cor} and n_{cla} are the effective refractive index, and Λ is the period of the LPG. For the uniform LPG there exist analytic solution for the equation (2). If only one cladding mode is excited, the un-uniform LPG of the equation can be solved by the transfer-matrix method. The LPG is assumed to be divided into N sections and each section is treat as a uniform LPG in index modulation and period, and then the transmission characteristics of the whole grating can be expressed as $T = \prod_{i=1}^{N} T_i$, where T_i is the transmission matrix of the section i.

$$T_i = \begin{bmatrix} \cos(\Omega_i dz) + i\frac{\delta_i}{\Omega_i}\sin(\Omega_i dz) & i\frac{k}{\Omega_i}\sin(\Omega_i dz) \\ i\frac{k}{\Omega_i}\sin(\Omega_i dz) & \cos(\Omega_i dz) - i\frac{\delta_i}{\Omega_i}\sin(\Omega_i dz) \end{bmatrix} \tag{5}$$

where $\Omega_i = \sqrt{k_i^{\,2} + \delta_i^{\,2}}$ **,** then the transmission of the whole LPG can be expressed as:

$$t = \frac{T_{22}}{(T_{22}T_{11} - T_{12}T_{21})} \tag{6}$$

In our problem, we first define an error function to evaluate the fitness of the particles (each particle is a set of the index modulations). The spectral window are divided into *m* discrete wavelength, the sum of the weighted errors is normally used as cost function.

$$Err(Particle_i) = \frac{1}{m}\sum_{l=1}^{m}\left(\frac{T_{target,l} - T_{i,l}}{\sigma_l}\right)^2, i = 1,2\ldots N \tag{7}$$

where the $Err(Particle_i)$ is the deviation of the calculated spectrum of particle i from the target spectrum, $T_{target,l}$ is the target spectrum component at the sampling point l, $T_{i,l}$ is the calculated spectrum component of particle i at the sampling point l, and σ_l is the uncertainty at the sampling point l. At the parts where the spectrum should be fitted better, the uncertainties should be smaller, to obtain good performance. In the simulation, we have to specify the upper and lower bounds of the grating physical parameters that should be optimized. The *N* particles are randomly initialized and each particle is a set of the grating index modulation parameters. In the upper and lower bonds the parameters values are continuous. The parameters of PSO algorithm is set as: ϕ_1=1.1, $\phi_2 = 3.0$, $N = 20$, $l = 500$, and w is tunable parameter expressed as:

$$w(iter) = w_{\max} - (iter / iter_{\max})(w_{\max} - w_{\min}) \tag{8}$$

where $iter$ and $iter_{\max}$ are the current and the maximum iteration number respectively. The advantage of tunable inertia weight are as follows: At the beginning stage, the inertia weight can be set a large value, so we can expand the searching space, when the potential particle has been approached to the best solution, the inertia weight must be very small, which make the PSO can implement elaborate searching around the best solution and avoiding escaping from the best solution to bad solutions. To improve the convergence solution to be the best results, we used local PSO (LPSO) where a special neighbor topology is introduced [12]. This can be seen in Fig.1, which shows two possible topologies used in our simulation.

4 Numerical Results and Discussions

In this section, two examples are demonstrated to test the effectiveness of the PSO algorithm for complex optimization designing problems in the fields of optical devices.

Mini Das proposed an improved version to realize wavelength multiplexing isolation filter using concatenated chirped LPG [14]. Although the introducing of chirp increased the useful channels by extending the loss window of the single LPG, however, to get high isolation between the channels, the key factor is to design the flat loss spectrum of the single LPG approach to 50%. The loss spectrum of the chirped LPG is approximately to be parabolic curve, which limited the useful channels, otherwise, in the chirp-type cascaded filter, the isolation undulation among the channels are very large. To overcome these problems, the best method is to design a wide bandwidth 3dss LPG filter within the bandwidth of 100nm, and the characteristics of the LPG pair interferometer made of two identical designed LPGB LPG filter. In the first example, we designed a 3dB loss sandwiched a standard fiber are also discussed.

The target spectrum loss in the region of 1500nm to1600nm should be 3dB, and outer the bandwidth the spectrum loss is 0. In our problem we mainly focus on the 3db loss bandwidth range. Fig.2 shows the optimized index modulations envelop using the improved LPSO method. In our simulation, we divided the fibre grating into

80 sections of the same length and assumed the index modulation in each section is unchanged; the whole length of the LPG is 5cm, the center resonance wavelength is 1550nm, the grating period is uniform along the whole LPG, the effective index difference between the core mode and cladding mode is 0.01. As can be seen, the designed index modulation profile is just like a sinc function, and the envelope is approximately symmetry. The maximum index modulation position is located in the center of the LPG, and the index modulation amplitude is less than 2×10^{-4}, so it is easy to realize such device for the low index modulation. Fig.3 (a) shows the output spectrum of the core mode of the single stage LPG calculated by the couple mode equation using the transfer matrix method. As can be seen, within the 100nm bandwidth that starts from 1500nm to 1600nm, the designed spectrums agree well with the target spectrums. There is a little divergence around the side band where the target transmission loss spectrum is a steep step, and break the continuity fortunately, however, those parts are not useful to our problem, and we are not concern about those region. Fig.3 (b) shows the divergence degree of the designed spectrum from the target spectrum within part of the useful bandwidth, we can see the divergence ripple is less than ±0.0002, the influence of the divergence on the transmission to the LPG pair interferometer can be ignored.

Fig.4 (a) shows the interference spectrum of the two identical optimized LPGs pair interferometer that made by cascading two LPGs together, and length of the sandwiched fibre between them is 20cm. We can see, in the bandwidth scope of more than100nm, the isolation degree at the stop bands is larger than 30db. What's more, the isolation undulation among channels is very small, and almost all the channels are in equal operation in the 100nm bandwidth region. This is because of the little divergence between the designed spectrum and that of the idea 3dB transmission loss filter.

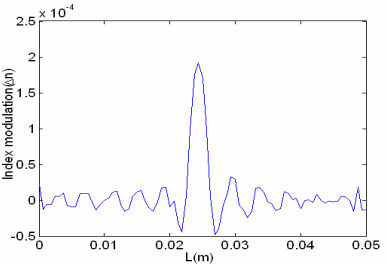


Fig. 2. The optimized index modulation envelop for 3dB LPG transmission spectrum

The channel spacing is about 1nm, so there are more than 100 useful equal spaced channels produced by the LPGs pair. To give a clear picture, Fig.4(b) shows part transmission spectrum of Fig.4 (a), in the central part, the isolation degree is about 40dB. The channel spacing can be tuned by changing the length of sandwiched fibre. With increasing the length of sandwiched fibre, the channel spacing will become narrower and we will get more useful channels. Fig.5 shows part of the delay spectrum, as can be seen, the dispersion effect of each pass band can be ignored, this is because the produced scheme of the channels are by interference, and the devices are

operated at transmission not reflection style. For single LPG, the delay difference acquired due to the group velocity difference between the core mode and cladding mode can cause the dispersion effect. But, in the devices discussed above, the whole length of the grating and the sandwiched fibre is very small, resulting a negligible dispersion. These characteristics make the device are very potential as interference filter in wave division multiplexing (WDM) systems.

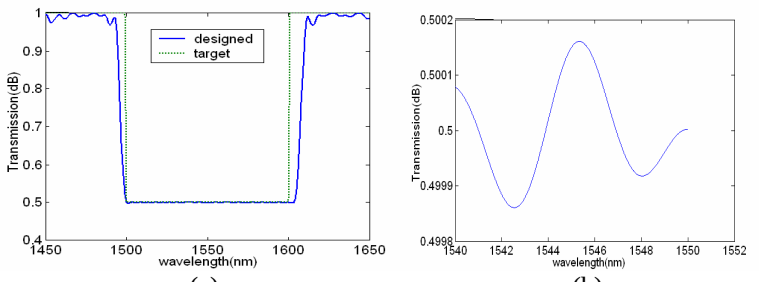


Fig. 3. The designed (solid line) and target (dotted line) transmission spectrum for the cladding mode (a) , The transmission spectrum ripples of the designed LPG(b)

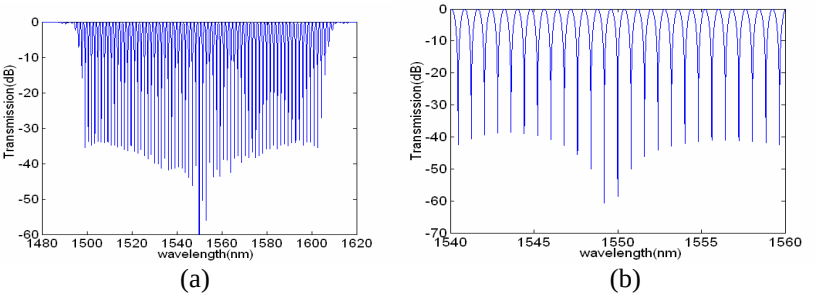


Fig. 4. The interference spectrum of the LPGs pair with inserted single mode fibre is 20cm (a), (b) is part of the transmission spectrum of (a)

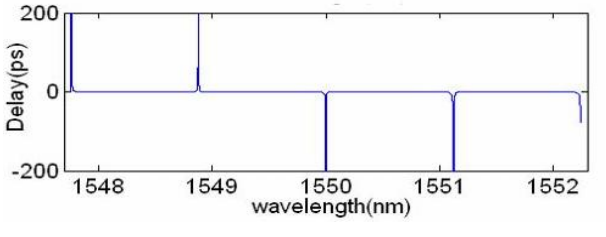


Fig. 5. The delay spectrum of the LPGs pair (only show part of the whole delay spectrum)

5 Conclusions

In conclusion, we proposed a novel synthesized method based on the improved intelligent LPSO algorithm. To make the algorithm convergence to a global optimal or a better sub-optimal, a special topology is used in our problem. An example to design a flatten loss spectrum in a large bandwidth that approach 50% is demonstrated to test the effectiveness of the algorithm. From the optimized LPG's sinc-type index modulation profile, we obtained a flat transmission spectrum in bandwidth of 100nm with the transmission loss approximately to be 50%. A numerical simulation showed the performance of the cascaded isolation filter consisted of two identical LPGs. A large number high isolation channels uniformly distributed in about 100nm bandwidth, so this type of filter is very suitable for used in the WDM systems. Compared with the chirped LPGs used for isolation filter, the LPG optimized by our method is easy to realize because there no chirp introduced. Otherwise, the isolation undulation is far less than cascading of two chirped LPGs. Based on the studied results, we believe that the LPSO algorithm is an effective inveiglement algorithm for optimally designing complicated LPG and other fibre grating filters. Since the LPSO is a stochastic search approach in nature, the required computation time cannot be precisely predicted. However, the advantage computer hardware can remedy this disadvantage. Otherwise, we may change some PSO control parameters to accelerate the convergence speed.

Acknowledgment

This work was supported by the National "973" Basic Research and Development Program of China (No.2003CB314901), the National "863" High Technology Project of China (No. No.2003AA311070).

References

1. Xuewen Shu, Lin Zhang and Ian Bennion, "Sensitivity characteristics of long period fiber gratings," IEEE J. Lightwave Technology, vol.20,no.2, pp.255-266.
2. K. W. Yang, A. G. Liu, Chih-Chun Cheng and Yu-Lung Lo, "Topology and shape optimizations of substrates for chirp fiber bragg grating spectrum tuning," IEEE J. Lightwave Technology, vol.20, no.7, pp.1182-1186, 2002.
3. Ashiish M, Vengsarkar, Paul J. Lemaire, Justin B. Judkins, Vikram Bhatia, Turan Erdogan, and John E. Sipe, "Long period fiber gratings as band-rejection filters," IEEE J. Lithrwave Technology, vol.14, no.1, pp.58-64, 1996
4. Donghui Zhao, Xuewen Shu, Yicheng Lai and Ian Bennion, "Fiber Bragg grating sensor interrogation using chirped fiber grating-based sagnac loop," IEEE J.Sensors, vol.3, no.6, pp734-738.
5. Bai-ou Guang, A-Ping Zhang, Hwa-Yaw Tam, Helen L.W.Chan, Chung-LoongChoy, "Step changed long-period fiber gratings," IEEE Photonics Technology Lett., vol.14, no.5, pp.657-659, 2002.3c
6. L. Wang and T. Erdogan, "Layer peeling algorithm for reconstruction of long- period fiber gratings," IEEE Electronics Lett., vol.37, no.3, pp.154-156, 2001

7. Johannes Skaar, Ligang Wang and Turan Erdogan, “On the synthesis of fiber Bragg gratings by layer peeling,” IEEE Lightwave Technology , vol.37, no.2, pp.165-173, 2001.
8. C. L. Lee and Y. Lai, “Evolutionary programming synthesis of optimal long period fiber grating filters for EDFA gain flattening,” IEEE photonics Technology Lett., vol.14, no.11, pp.1557-1559, 2003
9. Johannes Skaar and knut Magne Risvik, “ A genetic algorithm for the Inverse problem in synthesis of fiber gratings,” IEEE J. Lightwave Technology, vol.16, no.10, pp.1928-1932, 1998.
10. Carlos A. Coello, Gregorio Toscano Pulido and Maximino Salazar Lechuga, “Handing Multiple objectives with particle swarm optimization,” IEEE Transcations on Evolutionary computations, vol.8, no.3, pp.256-279, 2004.
11. Yumin Liu, Zhongyuan Yu, Hongbo Yang, Na Zhang, Qiang Feng and Xiaoguang Zhang, “Numerical optimization and simulation to wavelength-division multiplexing isolation filter consisted of two identical long period fiber grating,” Optics Communications 246, pp.367-372, 2005
12. Rui Mendes, James Kennedy and Jose Neves, “The Fully informed particle swarm: simpler, maybe better”, IEEE Transactions On Evolutionary computation, vol.8, no.3, pp.204-210, 2004
13. Yumin Liu, Zhongyuan Yu, Jianzhong Zhang, Bojun Yang, and Xiaoguang Zhang, “Optimization design of flat-band long-period grating,” Chinese Optics Letters, vol.2, no.3, pp: 200-202, 2004
14. Mini Das and Krishna Thyagarajan, “Wavelength -division multiplexing isolation filter using concatenated chirped long period gratings,” J. Optics Communications, 197, pp. 67-71, 2001,

The Kalman Particle Swarm Optimization Algorithm and Its Application in Soft-Sensor of Acrylonitrile Yield

Yufa Xu[1,2], Guochu Chen[1,2], and Jinshou Yu[1]

[1] Research Institute of Automation, East China University of Science and Technology, Shanghai 200237, China
[2] College of Electrical Engineering, Shanghai DianJi University, Shanghai 200240, China
xyf690@21cn.com

Abstract. This paper proposes kalman particle swarm optimization algorithm (KPSO), which combines kalman filter with PSO. Comparison of optimization performance between KPSO and PSO with three test functions shows that KPSO has better optimization performance than PSO. The combination of KPSO and ANN is also introduced (KPSONN). Then, KPSONN is applied to construct a soft-sensor of acrylonitrile yield. After comparing with practical industrial data, the result shows that KPSONN is feasible and effective in soft-sensor of acrylonitrile yield.

1 Introduction

Particle swarm optimization (PSO) is a stochastic optimization technique developed by Dr. Eberhart and Dr. Kennedy in 1995 [1], inspired by social behavior of bird flocking or fish schooling. PSO has been successfully applied in many areas: function optimization, neural network training, fuzzy system, and other areas where GA can be applied [2]. But, PSO suffers from an ineffective exploration strategy, especially around local optimal [3], thus can't find best solution as quickly as it can.

This paper presents an approach to particle motion that significantly speeds the search for global optimal fitness while improving on the premature convergence problems. The algorithm presented here, kalman particle swarm optimization algorithm (KPSO), bases its particle motion on kalman filtering and prediction.

2 The Kalman Particle Swarm Optimization Algorithm

The basic PSO assumes that each particle is a data structure that keeps track of its current position x and its current velocity v. Additionally, each particle memorizes the "best" position (best fitness) it has obtained in the past, denoted as p. The index of the global best particle is denoted as g. Each iteration step, the particle updates its position and velocity according to the following equations [1]:

$$v_{t+1} = \omega \cdot v_t + c_1 \cdot rand() \cdot (p - x_t) + c_2 \cdot Rand() \cdot (g - x_t) \quad (1)$$

$$x_{t+1} = x_t + v_{t+1} \quad (2)$$

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 176 – 179, 2006.

Where, ω is inertia weight. c_1 and c_2 are acceleration constants. $rand()$ and $Rand()$ are random numbers between $[0,1]$.

Kalman filters involve taking noisy observations over time and using model information to estimate the true state of the environment [4]. The kalman filter is limited to normal noise distributions and linear transition and sensor functions and is therefore completely described by several constant matrices and vectors. Specifically, given an observation vector Z_{t+1}, the kalman filter is used to generate a normal distribution over a belief about the true state. The parameters m_{t+1} and V_{t+1} of this multivariate distribution are determined by the following equations:

$$m_{t+1} = Fm_t + K_{t+1}(Z_{t+1} - HFm_t) \tag{3}$$

$$V_{t+1} = (I - K_{t+1})(FV_tF^T + V_X) \tag{4}$$

$$K_{t+1} = (FV_tF^T + V_X)H^T(H(FV_tF^T + V_X)H^T + V_Z)^{-1} \tag{5}$$

Where, F and V_x describe the system transition model while H and V_z describe the sensor model. The equations require a starting point for the filtered belief, represented by a normal distribution with parameters m_0 and V_0, which must be provided. The filtered or "true" state is then represented by a distribution:

$$x_t \sim Normal(m_t, V_t) \tag{6}$$

Here, the mean m_t is sampled once to obtain the value. After describing how to construct kalman filter, yielding m_t from Z_t, a simple form of prediction involves applying the transition model to obtain a belief about the next state m'_{t+1}:

$$m'_{t+1} = Fm_t \tag{7}$$

KPSO defines particle's motion entirely according to kalman prediction. Each particle keeps track of its m_t, V_t, and K_t. The particle then generates an observation for the kalman filter with the following equations:

$$z_v = \phi(g - x); \qquad z_p = x + z_v; \qquad Z = (z_p{}^T, z_v{}^T) \tag{8}$$

Where, ϕ is drawn uniformly from [0, 2.05]. The observation is then used to generate m_t and V_t according to equations (3), (4), and (5). Once the filtered value is obtained, a prediction m'_{t+1} is generated using equation (7), then,

$$X_{t+1} \sim Normal\ (m'_{t+2}, V_{t+1}) \tag{9}$$

The new state of the particle is obtained by sampling once from this distribution. The position of the particle may be obtained from the first half of x_{t+1}.

KPSO algorithm can be summarized in the following steps:

Step1: Initialize the parameters of KPSO and the position of each particle.
Step2: Evaluate the fitness of each particle according to the objective function.

Step3: Update the parameters of system according to equations (3)-(8).
Step4: Update the new position of each particle according to equation (9).
Step5: Update the fitness of each particle according to the objective function.
Step6: If necessary, update and store the global best particle of whole swarm.
Step7: If the stopping condition is not satisfied, go to step3. Otherwise, stop and obtain the result from the global best position and the global best fitness.

3 Experiment and Discussion

KPSO was compared with PSO to three test functions: Sphere, DejongF4, and Griewank. The three test functions are described as follow:

$$Sphere\ (x) = \sum_{i=1}^{d} x_i^2, \quad -50 \le x_i \le 50 \tag{10}$$

$$DeJongF\ \ 4(x) = \sum_{i=1}^{d} i x_i^4, \quad -20 \le x_i \le 20 \tag{11}$$

$$Griewank\ (x) = \frac{1}{4000}\sum_{i}^{d} x_i^2 - \prod_{i}^{d} \cos(\frac{x_i}{\sqrt{i}}) + 1, \quad -600 \le x_i \le 600 \tag{12}$$

The success rate and average convergence step are very important for any optimization algorithm. In order to compare the optimization performance of KPSO and PSO, the test experiment is designed as follows: The maximum value of iteration step is set to 2000, ω decay from 1.8 to 0.06, c_1, c_2 are set to 2, the population of KPSO and PSO are set to 30, the error limit for fitness is 10^{-5}. For every test function, there are 100 times of independent tests to be carried out. Then, test results are statistical accounted, averaged and summarized in Table 1.

Table 1. Optimization results for the three test functions with KSPSO and PSO

Functions	Algorithm	Ave. fitness	Best fitness	Success rate, %	Ave. step
Sphere	KPSO	0.0000	0.0000	100	863.7
	PSO	1.8572	0.0000	83	1812.4
DeJongF4	KPSO	0.0000	0.0000	100	507.8
	PSO	5.3389	0.0000	63	1744.6
Griewank	KPSO	0.1220	0.0000	94	713.1
	PSO	26.1992	0.0000	53	1800.1

It is clear from Table 1 that KPSO can find the global best fitness more easily and more quickly than PSO and KPSO can obtain better fitness than PSO. These results show that KPSO has better optimization performance than PSO.

4 KPSO's Application in Soft-Sensor of Acrylonitrile Yield

In process of acrylonitrile, acrylonitrile yield is one important guideline. It is important and useful for chemical plants to get the acrylonitrile yield. KPSO is applied to

train NN to construct neural networks called KPSONN. Here, KPSONN is used to construct a soft-sensor of acrylonitrile yield. The soft-sensor chooses six input signals that include pressure of reacting, temperature in middle unit, quantity of pure acrylonitrie, air percent, ammonia percent, and velocity of reacting. The output signal of the soft-sensor is acrylonitrile yield. This paper have trained the soft-sensing model by 267 groups of training sample data which is processed by error-detection, smoothing, normalization. After model learning and accounting, average variance is 0.342; average value of error is 0.221. This result shows that the fitness precision between the model's predictive values and real values is high. Then the soft-sensing model is examined by another 50 groups of examining sample data. The comparison between predictive values and real values is showed in Fig. 1.

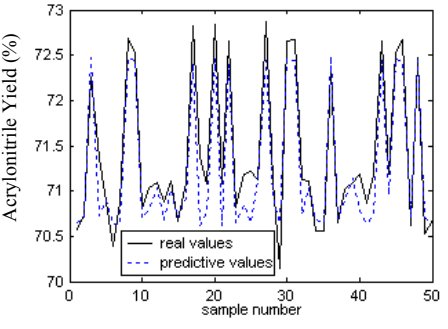


Fig. 1. The comparison between predictive values with KPSONN and real values

In 50 groups of examining sample data, average variance is 0.198; average value of error is 0.160. These data and Fig. 1 show that the predictive precision between the model's predictive values and real values is high.

5 Conclusion

KPSO is a new improved PSO algorithm. KPSO can find the global best solution effectively, and has better performance than PSO. KPSO is feasible and effective in soft-sensor of acrylonitrile yield.

References

1. Kennedy J, Eberhart R. Particle Swarm Optimization. Proc IEEE Int. Conf on Neural Networks. Perth, 1995: 1942-1948
2. Eberhart R C, Shi Y. Particle swarm optimization: developments, applications and resources. Proc. 2001 Congress on Evolutionary Computation. Soul, South Korea, 2001: 81-86
3. Bergh F. van den, Engelbrecht A. P.. A study of particle swarm optimization particle trajectories. Information Sciences. 176 (2006): 937-971
4. Russel. S., Norvig. P.. Artificial Intelligence: A Modern Approach. Second Prentice Hall, Englewood Cliffs, New Jersey, 2003

Data Fitting Via Chaotic Ant Swarm

Yu-Ying Li[1,2], Li-Xiang Li[3], Qiao-Yan Wen[1], and Yi-Xian Yang[3]

[1] School of Science, Beijing University of Posts and Telecommunications, Beijing, 100876, China
liyuying19770630@sohu.com
[2] State Key Laboratory of Integrated Services Network, Xidian University, Xi'an, 710071, China
[3] Information Security Center, Department of Information Engineering, Beijing University of Posts and Telecommunications, Beijing, 100876, China

Abstract. A novel method of data fitting via chaotic ant swarm (CAS) is presented in this paper. Through the construction of a suitable function, the problem of data fitting can be viewed as that of parameter optimization, and then the CAS is used to search the parameter space so as to find the optimal estimations of the system parameters. To investigate the performances of the CAS, the CAS is compared with the particle swarm optimization (PSO) on two test problems. Simulation results indicate that the CAS achieves better performances.

1 Introduction

Data fitting is one of the crucial steps to analyze the data produced in experimental, social, and behavioral science, because it can not only explain the data but also do some predictions for the future. Data fitting includes three important steps: Firstly choose a suitable function flexibly from a linear, polynomial, exponential function or any other function in accordance with the changing tendency of the data; Secondly determine the parameters of the chosen function by using some kind of method; Finally make the sum of square errors tend to zero infinitely. Thus, we can convert the problem of data fitting to that of parameter optimization of the chosen function.

Many data fitting techniques have been proposed recently, such as data fitting by spline function, exponential data fitting. In this paper we present a novel method based on the CAS [1,2] to solve the problem of data fitting.

The remainder of this paper is organized as follows. Section 2 reviews the CAS. Section 3 describes the parameters estimation process of data fitting via the CAS. Section 4 compares the CAS with the PSO [3] on two problems of data fitting and gives the discussions. Finally, a summary is given in section 5.

2 Chaotic Ant Swarm Optimization

By connecting the chaotic behavior of individual ant with the intelligent optimization actions of ant colony, a novel optimization method, the CAS, was

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 180–183, 2006.

developed. To explain their connection, the organization variable is introduced to adjust the chaotic behavior of individual ant and lead the individual to move to the new position acquired the best fitness eventually. In this paper we use the global version of the CAS in which the whole swarm is considered as the neighborhood. In this version the ants are manipulated according to (1):

$$\begin{cases} y_i(t) &= y_i(t-1)^{(1+r_i)} \\ z_{id}(t) &= (z_{id}(t-1) + \frac{7.5}{\psi_d} \times v_i) \\ & \quad \times exp((1-exp(-ay_i(t)))(3-\psi_d(z_{id}(t-1) + \frac{7.5}{\psi_d} \times v_i))) \\ & \quad -\frac{7.5}{\psi_d} \times v_i + exp(-2ay_i(t)+b)(p_{id}(t-1) - z_{id}(t-1)) \,. \end{cases} \tag{1}$$

where $i = 1, 2, \cdots, N$, N is the swarm's size. t means the current time step. $y_i(t)$ is the current organization variable of ant i, $y_i(0) = 0.999$. $r_i (0 < r_i < 1)$ is the organization factor of ant i. $\psi_d (\psi_d > 0)$ determines the selection of the search range of dth element of variable in search space . If the interval of the search is $[0, \omega_d]$, we can obtain $\omega_d \approx \frac{7.5}{\psi_d}$. $v_i (0 < v_i < 1)$ determines the search region of ant i. a is a sufficiently large positive constant. $b (0 \leq b \leq \frac{2}{3})$ is a constant. $p_{id}(t-1)$ is the best position found by the ith ant and its $N-1$ neighbors within $t-1$ steps. $z_{id}(t)$ is the current state of the dth dimension of ant i, and $z_{id}(0) = \frac{7.5}{\psi_d} \times (1 - v_i) \times rand(1)$ where $rand(1)$ is a random number uniformly distributed within the range $[0, 1]$.

Equation (1) describes the search process of the CAS. The organization variable is used to control the chaotic process of ant moving, and its influence on the ant's behavior is very weak initially. That is, initially the organization capabilities of the ants are very weak so that a non-coordinated process occurs which is characterized by the chaotic walking of ants. This phase lasts until the influence of organization on the individual behavior is sufficiently large. Then, the chaotic behavior of the individual ant disappears and a coordination phase starts. That is, ants do some further searches and move to the best position which they have ever found in search space. Throughout the whole process, these ants exchange information with other ants, then compare and memorize the information.

3 Parameters Identification of Data Fitting Via the CAS

In this paper, the chosen function is $y = f(x, c)$, where $x = (x_1, \cdots, x_M)^T$ is the state vector, and $c = (c_1, c_2, \cdots, c_l)^T$ is the unknown parameter vector. The fitness function was defined: $F = \sum_{k=1}^{M} [y_k - f(x_k, c)]^2$, where (x_k, y_k) is one known point of the data and M is the number of the known points. The parameter identification process of data fitting via the CAS is expressed as follows:

1. Randomly generate the organization variable factors and the initial positions of all ants in the respective search space, where the initial position of the ith ant is $\hat{c_i} = (\hat{c_{i1}}, \cdots, \hat{c_{il}})^T$ and $\hat{c_i}$ is the estimation of c;

2. Compute fitness values of N ants according to the fitness function, and then find the minimal fitness value F_{min} among N ants and its coordinates p;

3. Change organization variables and current states of N ants according to (1), and then find the current minimal fitness value $F_{min}{}'$ among N ants and its coordinates p'. If $F_{min}{}'<F_{min}$, we assign $F_{min}{}'$ to F_{min}, and assign p' to p;

4. Repeat step 3 until a given maximal number of iterations is achieved.

4 Experimental Results

In our experimental study, the CAS was compared with the PSO to investigate the performances of the CAS by two test problems [4] of data fitting. The two test problems were defined in Table 1, and their figures were reported in Fig.1. According to the changeful tendency of the points in Fig.1, the chosen function is $f(x) = c_1x + c_2x^2 + \cdots + c_lx^l$ for Test problem 1 while the chosen function was $f(x) = c_4/(c_1 + e^{c_2x+c_3})$ for Test problem 2. For each test problem, the parameters of the CAS were $a = 200$, $b = \frac{7}{12}$, $\psi_d = 0.75$, $r_i = 0.05 + 0.1 \times rand(1)$, $v_i = rand(1)$, and the parameters of the PSO were $c_1 = c_2 = 0.5$, ω was gradually decreased from 1.2 toward 0.4 during the three quarters of the maximum allowed number of iterations. For each test problem using the CAS and the PSO, the maximum number of allowed function evalutions was set to 500, and 50 experiments were performed, starting from the range $[-10, 10]^l$, where l was the dimensions of the corresponding problem. By the search of the CAS, the best function is $f(x) = 2.9972x - 4.7480x^2 + 2.1553x^3 + 3.1261x^4 - 2.5478x^5$ for Test problem 1 while $f(x) = 3.5523/(3.5666 + e^{-2.0325x+1.2574})$ for Test problem 2. The search results of the CAS and the PSO are reported in Table 2. From Table 2, the CAS achieves better performances.

Table 1. Starting with the following set of data for each test problem

Test problem 1	X	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
	Y	0.00	0.30	0.40	0.55	0.63	0.71	0.77	0.84	0.89	0.95	1.00
Test problem 2	X	-2.0	-1.6	-1.2	-0.8	-0.4	0.0	0.4	0.8	1.2	1.6	2.0
	Y	0.018	0.039	0.083	0.168	0.310	0.500	0.690	0.832	0.917	0.961	0.982

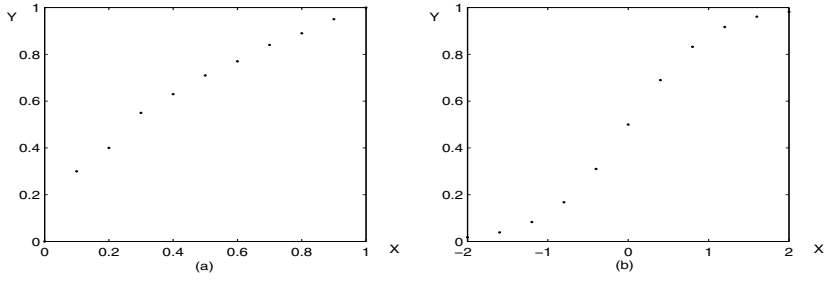


Fig. 1. (a) is the simulation for Test problem 1 while (b) is for Test problem 2

Table 2. The experimental results for Test problem 1 and Test problem 2

Test problem	Test problem 1						Test problem 2	
experiment method	CAS	PSO	CAS	PSO	CAS	PSO	CAS	PSO
dimension number	3	3	4	4	5	5	4	4
best fitness value	0.0059	0.0060	0.0059	0.0063	0.0043	0.0046	0.0015	0.0051
the mean	0.0222	0.0279	0.0508	0.0512	0.0375	0.0382	0.0016	1.3999

5 Conclusions

This paper not only presents a novel method based on the CAS to cope with the problem of data fitting, but also investigates its performances to solve the problem of data fitting by comparing with the PSO. Experimental results show the CAS achieves better performances. In conclusion, the CAS is an efficient alternative for solving the problem of data fitting.

Acknowledgement

This work is supported by the Major Research Plan of the National Natural Science Foundation of China, Grant No. 90604023; the National Natural Science Foundation of China, Grants No. 60373059, 90204017, 60372094; the National Research Foundation for the Doctoral Program of Higher Education of China, Grants No.20040013007; the ISN Open Foundation; Key Project of Chinese Ministry of Education, Grant No. 205033; the National 973 of China, Grant No. TG1999035804; the Natural Science Foundation of Beijing, Grant No. 4062025.

References

1. L. Li, H. Peng, X. Wang, Y. Yang, An Optimization Method Inspired by "Chaotic" Ant Behavior, International Journal of Bifurcation and Chaos 16, 2006. (Appear in July 2006)
2. L. Li, Y. Yang, H. Peng, X. Wang, Parameters Identification of Chaotic Systems via Chaotic Ant Swarm, Journal of Chaos Solitons Fractals 28 (2006) 1204-1211.
3. K. Parsopoulos, M. Vrahatis, Recent Approaches to Global Optimization Problems through Particle Swarm Optimization, Natural Computing 1 (2002) 282-288.
4. H. A. Kamal, M. H. Eassa, Solving Curve Fitting Problems Using Genetic Programming, Proc. 11th IEEE Mediterranean Conference on Electrotechnic, Cairo, Egypt, (2002) 316-321.

A Hybrid Discrete Particle Swarm Algorithm for Hard Binary CSPs

Qingyun Yang[1,2], Jigui Sun[1,2,3], Juyang Zhang[1,2], and Chunjie Wang[4]

[1] College of Computer Science and Technology, Jilin University, Changchun, 130012, China
[2] Key Laboratory for Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, 130012, China
[3] Open Laboratory for Intelligence Information Processing, Fudan University, Shanghai, 200433, China
[4] Basic Sciences of ChangChun University of Technology, ChangChun University of Technology, Changchun, 130012, China
qyyang_jlu@yahoo.com.cn, jgsun@jlu.edu.cn, sl_zj@vip.sina.com, chunjie_wang@yahoo.com.cn

Abstract. The discrete particle swarm algorithm for binary constraint satisfaction problems (CSPs) is analyzed in this paper. The analysis denotes that φ_1 and φ_2 are set to 0 may be a heuristic similar to min-conflict heuristic. The further observation is the impact of local best positions. A control parameter p_b is introduced to reduce the effect of the local best positions. To improve the performance, simulated annealing algorithm is combined with the discrete particle swarm algorithm, and the neighborhood exploring in simulated annealing is carried out by ERA model. Eliminating repeated particles and Tabu list avoiding cycling are also introduced in this paper. Our hybrid algorithm is tested with random constraint satisfaction problem instances based on phase transition theory. The experimental results indicate that our hybrid discrete particle swarm algorithm is able to solve hard binary CSPs.

1 Introduction

Particle Swarm Optimization (PSO) is a stochastic search technique developed by Kennedy and Eberhart[1], which is inspired by the behavior of bird flocking and fish schooling. PSO exploits a population of potential solutions to probe the search space. Each particle is assigned a randomized velocity and is iteratively moved through the search space. PSO is widely used as optimizer for continuous nonlinear functions. Many experimental results denote that PSO is a excellent stochastic search method on continuous numeric optimization problems. Basically the PSO is developed to solve non-constrained optimization problems. But recently PSO is proposed to solve constrained optimization problems[2],[3] and integer programming[4]. Constraint satisfaction problems are usually discrete and are not so easily solved by PSO. [5] proposed a discrete particle swarm algorithm to solve binary constraint satisfaction problems (we call it PS-CSP),

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 184–193, 2006.

which is a general version of Clerc's[6]. The experimental results indicate the discrete PSO is competitive with other algorithms such as HC, ant colony[7].

The remainder of this paper is organized as follows: Section 2 provides a short overview of constraint satisfaction problem and the discrete particle swarm algorithm PS-CSP, outlining the basic concept of CSP and the basic operations redefined in PS-CSP. PS-CSP is analyzed in Section 3 and a substituted version is proposed to improve the performance. In Section 4 the simulated annealing is introduced to enhance the global best particle, the neighborhood particle in simulated annealing is generated by ERA model. In Section 5 we add nohope to increase the opportunity for searching and add tabu list to prevent cycle in the search procedure. Some experimental results are given in Section 6. Section 7 is the conclusion.

2 Preliminaries

A Constraint Satisfaction Problem (CSP) is a triplet (X, D, C), here $X = \{X_1, X_2, \ldots, X_n\}$ is a set of Variables, which may take on values from a set of domains $D = \{D_1, D_2, \ldots, D_n\}$, and a set of constraints $C = \{C_1, C_2, \ldots, C_m\}$. A constraint C_i on the ordered set of variables $X(C_i) = (X_1, \ldots, X_n)$ is a subset of the Cartesian product $D(X_1) \times \ldots \times D(X_n)$ that denotes the compatible pairs of value for variable $X_1, \ldots, X_n$. A binary constraint C_{ij} on variable X_i and X_j is a set of pairs, C_{ij} allows for X_i to take the value v_i and X_j to take the value v_j iff $(v_i, v_j) \in C_{ij}$ and we say the binary constraint is *satisfied* otherwise it is *violated*. We call a CSP binary constraint satisfaction problem iff all the constraints in CSP are binary constraints. Conflict number of x is the number of constraints who contain the variable x and violate the current assignment, $conflict(x) = \sum_{i=1}^{|C|} violate(c_i)$. A solution of a constraint satisfaction problem is an assignment of values to the set of X such that all the constraints are satisfied simultaneously.

PS-CSP is a discrete particle swarm algorithm based on the idea of[6],[8], which redefined the operators for velocity. PS-CSP throws all the particles into the problem space just the same as the basic particle swarm algorithm does at the initial phase. Velocity is a real number in basic particle swarm algorithm while the subtraction of two positions resulting in a velocity in PS-CSP. Supposing $\vec{x}$ and $\vec{y}$ are positions. Then $\vec{v} = \vec{x} \odot \vec{y} = \{\vec{y} \rightarrow \vec{x}\}$ denotes a velocity. The long arrow indicates the change of position. Additional operators need to be redefined for the changing of velocity. The addition of a position with a velocity results in a position, $\vec{x} \oplus \vec{v}$, where suppose $\vec{x}$ is a position and $\vec{v} = \vec{z} \odot \vec{y}$. Then $\vec{x} \oplus \vec{v}$ equals the position produced in by $\vec{x} \oplus \vec{v} = \begin{cases} z_i \; if \; x_i = y_i \\ x_i \; otherwise \end{cases}$. The next operator is a velocity add another velocity resulting in a new velocity $\vec{v} \circ \vec{w}$, where $\vec{v} = \vec{b} \odot \vec{a}$, $\vec{w} = \vec{y} \odot \vec{x}$. Then the newly created velocity is $\vec{v} \circ \vec{w} = \begin{cases} a_i \rightarrow y_i \; if \; b_i = x_i \\ a_i \rightarrow b_i \; otherwise \end{cases}$. The last operation is the multiplication of a velocity and a

coefficient, according to a corresponding position. The multiplication results in a velocity vector $\varphi \otimes \overrightarrow{v}$, where $\overrightarrow{v} = \overrightarrow{y} \odot \overrightarrow{x}$. The result velocity vector produced in by $\varphi \otimes \overrightarrow{v} = \begin{cases} x_i \to x_i \; if \; nbconf_i \leq \varphi \\ x_i \to y_i \; otherwise \end{cases}$, $nbconf_i$ is the conflict number of position $\overrightarrow{x}$ with offset i . The calculation of conflict number uses the equation defined above. At last the main formulations of the discrete particle swarm algorithm can be rewritten as follows:

$$\overrightarrow{v}^{\,t} = (\varphi_1 \otimes (\overrightarrow{p} \odot \overrightarrow{x}^{\,t-1})) \circ (\varphi_2 \otimes (\overrightarrow{g} \odot \overrightarrow{x}^{\,t-1})) \tag{1}$$

$$\overrightarrow{x}^{\,t} = \overrightarrow{x}^{\,t-1} \oplus \overrightarrow{v}^{\,t} \tag{2}$$

3 Analysis of the Discrete Particle Swarm

The formula (1) is partitioned into two parts in the algorithmic description of [5], i.e. $\overrightarrow{v}^{\,t} = v' \circ v''$, where $v' = \varphi_1 \otimes (\overrightarrow{p} \odot \overrightarrow{x}^{\,t-1})$ and $v'' = \varphi_2 \otimes (\overrightarrow{g} \odot \overrightarrow{x}^{\,t-1})$. With the definition of the operators in Section 2, we can see that v' can be regarded as $\left|x_{ij}^{t-1} \to \begin{matrix} bestSoFar_{ij} & if & nbConf > \varphi_1 \\ x_{ij}^{t-1} & if & nbConf \leq \varphi_1 \end{matrix}\right|$ and v'' can be $\left|x_{ij}^{t-1} \to \begin{matrix} Rand & if & nbConf > \varphi_2 & and & deflection \\ globalBest_j & if & nbConf > \varphi_2 & and & not & deflection \\ x_{ij}^{t-1} & if & nbConf \leq \varphi_2 \end{matrix}\right|$, where $nbConf$ is the conflict number of variable x_j, $deflection$ is a sufficient small positive real value of the threshold for random selection. Then we get the addition of two velocities with v' and v''. We have the formula $\overrightarrow{v}^{\,t} = v' \circ v'' = \left|x_{ij}^{t-1} \to \begin{matrix} bestSoFar_{ij} & if & nbConf > \varphi_1 \\ x_{ij}^{t-1} & if & nbConf \leq \varphi_1 \end{matrix}\right| \circ \left|x_{ij}^{t-1} \to \begin{matrix} Rand & if & nbConf > \varphi_2 & and & deflection \\ globalBest_j & if & nbConf > \varphi_2 & and & not & deflection \\ x_{ij}^{t-1} & if & nbConf \leq \varphi_2 \end{matrix}\right|$ and have

$$\overrightarrow{v}^{\,t} = \left|x_{ij}^{t-1} \to \begin{matrix} bestSoFar_{ij} & if & nbConf > \varphi_1 & and & x_{ij}^{t-1} \neq bestSoFar_{ij} \\ Rand & if & nbConf > \varphi_2 & and & deflection \\ globalBest_j & if & nbConf > \varphi_2 & and & not & deflection \\ x_{ij}^{t-1} & if & nbConf \leq \varphi_2 \end{matrix}\right| \tag{3}$$

After the update of the velocity at time (iteration) t with formula (3), we achieve the position at time t as follows $\overrightarrow{x}^{\,t} = \overrightarrow{x}^{\,t-1} \oplus \overrightarrow{v}^{\,t} =$

$$\overrightarrow{x}^{\,t-1} \oplus \left|x_{ij}^{t-1} \to \begin{matrix} bestSoFar_{ij} & if & nbConf > \varphi_1 & and & x_{ij}^{t-1} \neq bestSoFar_{ij} \\ Rand & if & nbConf > \varphi_2 & and & deflection \\ globalBest_j & if & nbConf > \varphi_2 & and & not & deflection \\ x_{ij}^{t-1} & if & nbConf \leq \varphi_2 \end{matrix}\right|,$$

i.e.

$$x_{ij}^{t} = \begin{cases} bestSoFar_{ij} & if \quad nbConf > \varphi_1 \quad and \quad x_{ij}^{t-1} \neq bestSoFar_{ij} \\ Rand & if \quad nbConf > \varphi_2 \quad and \quad deflection \\ globalBest_j & if \quad nbConf > \varphi_2 \quad and \quad not \quad deflection \\ x_{ij}^{t-1} & if \quad nbConf \leq \varphi_2 \end{cases} \tag{4}$$

We can see from formula (4) that φ_1 and φ_2 play important roles in the update of position throughout the search procedure. Their values determine whether the PS algorithm converges or not. PS-CSP tests both φ_1 and φ_2 being set to 0 and 1 respectively. Results indicate that better performance it can achieve when φ_1 and φ_2 are set 1 and the diversity is enhanced too. Higher diversity means more powerful global search ability. In fact if we consider the fourth element in formula (4) we can see that φ_2 being set to 1 may not as efficient as mentioned in PS-CSP.The position value of x_{ij} is reserved into the next iteration even if x_{ij} exists conflict (i.e. $nbConf = 1$) when φ_2 is set to 1. Thus the algorithm remains the conflicting value in the population and it does not converge until each the position value happens to have non-conflict. It is obvious that the non-conflict position value propagates from time (iteration) $t-1$ to t when φ_2 is set to 0. In fact φ_2 is set to 0 is a heuristic similar to min-conflict heuristic. The same situation can be applied to φ_1, i.e. we replace the current position value by $bestSoFar_{ij}$ only non-conflict exists. Although φ_1 and φ_2 be set to 0 reduces the diversity in the population, the experiments indicate they are more efficient, and we produce in an opposite result as did in PS-CSP. Thus in our algorithm we merge φ_1 and φ_2 into another parameter, namely φ. φ is set to 0 in our experiments. Now we get the particle swarm algorithm as follows:

$$\vec{v}^{\,t} = \varphi \otimes ((\vec{p} \odot \vec{x}^{\,t-1}) \circ (\vec{g} \odot \vec{x}^{\,t-1})) \tag{5}$$

$$\vec{x}^{\,t} = \vec{x}^{\,t-1} \oplus \vec{v}^{\,t} \tag{6}$$

The update of x_{ij} is refined as follows:

$$x_{ij}^{t} = \begin{cases} bestSoFar_{ij} & if \quad nbConf > \varphi \quad and \quad x_{ij}^{t-1} \neq bestSoFar_{ij} \\ Rand & if \quad nbConf > \varphi \quad and \quad deflection \\ globalBest_j & if \quad nbConf > \varphi \quad and \quad not \quad deflection \\ x_{ij}^{t-1} & if \quad nbConf \leq \varphi \end{cases} \tag{7}$$

$bestSoFar_{ij}$ is a local attractor. In the update of x_{ij}, the conflict is checked at first. If the conflict is larger than 0 then the equality of x_{ij} and $bestSoFar_{ij}$ is determined. If $x_{ij} \neq bestSoFar_{ij}$ then $bestSoFar_{ij}$ is assigned to x_{ij} directly and entered into the next iteration. In some certain extremely situation when $bestSoFar_{ij}$ unsatisfies the constraints and so does $globalBest_j$, the only way to satisfy $nbConf \leq \varphi$ is to select random value ($Rand$ in algorithm). But $deflection$ itself is a very small positive value (the quasi-optimal value is $2/n$), thus x_{ij} always wanders about the $bestSoFar_{ij}$ and $globalBest_j$, which may decrease the performance of the algorithm. We believe that the search ability

can be enhanced if the $bestSoFar_{ij}$ is reduced, and the particles converge to the global best particle is a good strategy. We add a control parameter $p_b \in [0,1]$ to the first part of the formula (7). The algorithm assigns $bestSoFar_{ij}$ to x_{ij} for entering the next iteration with probability p_b when $nbConf > \varphi \quad and \quad x_{ij} \neq bestSoFar_{ij}$. We define a operator $\bullet$ to denote the probability selection. At last the discrete particle swarm algorithm is as follows:

$$\vec{v}^{\,t} = \varphi \otimes ((deflection \bullet (Rand \odot \vec{x}^{\,t-1})) \circ (p_b \bullet (\vec{p} \odot \vec{x}^{\,t-1})) \circ (\vec{g} \odot \vec{x}^{\,t-1})) \quad (8)$$

$$\vec{x}^{\,t} = \vec{x}^{\,t-1} \oplus \vec{v}^{\,t} \quad (9)$$

Then the update of x_{ij} can be

$$x_{ij}^t = \begin{cases} bestSoFar_{ij} & if \quad nbConf > \varphi \quad and \quad x_{ij}^{t-1} \neq bestSoFar_{ij} \quad and \quad p_b \\ Rand & if \quad nbConf > \varphi \quad and \quad deflection \\ globalBest_j & if \quad nbConf > \varphi \quad and \quad not \quad deflection \\ x_{ij}^{t-1} & if \quad nbConf \leq \varphi \end{cases} \quad (10)$$

4 Combining Simulated Annealing with Discrete Particle Swarm

4.1 Simulated Annealing for Global Best Particle

Simulated annealing is an advanced local search method which, introduced by Kirkpatrick, Gelatt and Vecchi[9], inspires the physical annealing process studied in statistical mechanics. At each iteration, starting from an initial solution s, a new solution s' in the neighborhood of s is generated randomly. Then a decision is taken to decide whether s' will replace s based on the calculation of $\triangle = f(s') - f(s)$. For a minimization problem, if $\triangle \leq 0$, we move from s to s', otherwise, we move to s' with the probability $e^{-\frac{\triangle}{t}}$, where t is a control parameter called the temperature (higher temperatures lead to higher accepting probabilities and vice versa). Usually SA algorithm starts from a high temperature and decreases gradually. The algorithm will stop until the termination criterion is reached.

SA can enhance the ability of local search. Applying SA to all the particles is a straightforward way when we combine SA with particle swarm. But the computation cost increases if this method is adopted. On the consideration of the performance, we apply the SA procedure only to global best particle and propagate the results of SA to other particles with particle swarm algorithm.

4.2 Generating Neighbor with ERA Model

The exploring of the neighborhood structure may greatly influence the performance of SA. A good neighbor can increase the likelihood of finding better solution. Many methods can be used to generate neighbor.

ERA(Environment, Reactive rules, Agent) model[10] is similar to min-conflicts heuristic[11] though it is a distributed multi-agent system based on the theory of artificial life. [10] compared ERA model with min-conflicts heuristic and pointed out the ERA model is superior to min-conflicts. In ERA, each agent represents a variable in CSP. Agents live in lattice environment, and sense the local and global information around them. The environment records agents' current state and computes the constraint violations for each value in the domains of all variables. Each agent can only move within a row it located in and has predefined local reactive behaviors: *better_move*, *least_move* and *random_move*. Agents select them with some certain probability under the pressure of environment that based on constraint violations. *better_move* is to find a position with less violation number with a probability of *better_p*, *least_move* is to find the least violation number position with a probability of *least_p*, and *random_move* is to move randomly with a probability of *random_p*. *better_move* has less time complexity than *least_move* and may be more efficient at the early search stage. While obviously the *least_move* can accelerate convergence to local optima or global optima. *random_move* is a necessary technique to avoid getting stuck in local optima.

Three local behaviors have been analyzed in [10], and pointing out that maybe exist $random\_p < better\_p < least\_p$ for general purpose. Their experiments also indicate $least\_p \ / \ random\_p = \mathrm{n}$ (here n is the domain size of the specified problem) may be a good choice. We use this empirical setting for our test. From the algorithmic description of the ERA it can be concluded that there exists a approximate formulation, namely $better\_p + random\_p + least\_p = 1$. Thus if *better_p* is specified, we can compute *random_p* and *least_p* with empirical approximate formulations:

$$random\_p = (1 - better\_p)/(d + 1) \tag{11}$$

$$least\_p = random\_p \times d \tag{12}$$

where d is the domain size of a constraint satisfaction problem(assuming all variables have the same domain size).

5 Further Improvement

5.1 Eliminating Repeated Particles

As mentioned in PS-CSP, if φ_1 and φ_2 are set to 0, the Hamming distance will be lower. In our observation the same position takes place in the running both in PS-CSP and our newly developed substitute. It shrinks the population size indirectly and is helpless in search procedure. To enlarge the diversity among particles, we compare the Hamming distance between the current particle and the global best particle. If the distance is zero i.e. the two particles have the same position, we redistribute the current particle, however, remains the local best position. In order to keep the result that the current particle has achieved, we redistribute only partial elements in the position.

5.2 Avoiding Cycle with Tabu List

Stochastic search algorithm may repeatedly enter into the same status, i.e. particle swarm optimization find a position repeatedly. It is useless in the search procedure, especially after the nohope activity. We introduce a short tabu list to avoid cycling. Each iteration the global best position is recorded, and is compared with the elements in the tabu list. If the global best position already exists in the tabu list, a repair is activated by a more chaotic SA procedure(where we increase *random_p* in ERA model). When the global best particle arrives at a new position(i.e. a less fitness it approaches) we add it into the tabu list. a fist-in-fist-out queue is used for tabu list in our algorithm and only global best particle uses tabu list.

6 Experimental Results

To compare the performance of PS-CSP with the revised algorithm and our hybrid discrete particle swarm algorithm, first we choose the Model E[12] class with 15 variables and 15 domain elements as PS-CSP did, and choose JavaCsp (JavaCsp is a random binary constraint satisfaction problem instance generator)[1] to generate instances. Model E has one parameter named p. The generation process is to select uniformly, independently and with repetition, $pd^2n(n-1)/2$ conflicts out of the $d^2n(n-1)/2$ possible, where d is the domain size and the n is the number of variables. In our observation, the phase transition point is approximately 0.31 in the development of JavaCsp. Thus we set parameter p at 11 values equals spaced between 0.21 and 0.31 instead of between 0.2 and 0.38. Random constraint satisfaction problem theories proved these values located in the so called *mushy region*, where interesting instances with few solutions can be sampled. Each value of p we sample 5 instances; each instance has at least one solution. In our test all algorithms stop when they have found a solution or when it has performed 10^6 evaluations in each case. 10^6 evaluations are considered no more solution can be found and are excluded from experimental results. 10 runs are performed for each instance. The experiments are made on PC(Intel Celeron CPU 1.8GHz, 512M RAM, Window XP SP2 and JDK 1.5).

The parameters of PS-CSP in this test are set the quasi-optimal values described in [5]. Those values include: swarm size is 50, deflection = 2/n, where n is the number of variables, $\varphi_1 = \varphi_2 = 1$. the nohope times is set to 100. For our hybrid algorithms parameters, the swarm size, $deflection$ and nohope times reserve the same values as PS-CSP. $p_b = deflection/30$, and tabu list length is set to 15. The parameter setting for SA including: the initial temperature T0=5, the terminating criterion temperature tend=0.02, the epoch length L is set to the dimension of a particle (i.e. the number of variables in the constraint network), and decreasing rate $b = 0.95$. The parameter setting for ERA model has only one value to be tuned for experimental convenience, given by our experience, where *better move* = 0.15. In the repair phase we give more random

[1] http://www.xs4all.nl/∼bcraenen/JavaCsp/resources/javacsp_1.0.2.jar.

Table 1. Comparison results with the success rate and the mean iteration of Model E class

p	PS-CSP		PS-CSP$_{\varphi=0}$		PS-CSP$_{\varphi=0}, p_b$		HPS-CSP	
	succ.	mean iter.	succ.	mean iter.	succ.	mean iter.	succ.	mean iter.
0.21	50/50	350.7	50/50	68.9	50/50	54.5	50/50	0.7
0.22	50/50	621.1	50/50	198.1	50/50	99.4	50/50	0.6
0.23	50/50	1522.1	50/50	257.7	50/50	159.6	50/50	0.9
0.24	49/50	7978.8	50/50	1048.7	50/50	253.2	50/50	0.9
0.25	45/50	26260.4	50/50	5004.1	50/50	541.2	50/50	0.9
0.26	30/50	63732.6	31/50	14287	50/50	1733.3	50/50	1.6
0.27	27/50	59207.1	35/50	19022.2	50/50	2017.3	50/50	1.1
0.28	11/50	24638.9	20/50	39142.8	50/50	4111	50/50	4
0.29	8/50	183358.5	12/50	228299	50/50	2189.7	50/50	3
0.30	0/50	–	1/50	694652	50/50	72442.8	50/50	41.6
0.31	0/50	–	2/50	19892	50/50	32487.8	50/50	353.3

opportunity to the algorithm to get out of local optima. The parameter change to *better move* = 0.1.

From the table 1 we can see that PS-CSP has the worst performance comparing with other algorithms. PS-CSP use the quasi-optimal parameter setting with $\varphi_1 = \varphi_2 = 1$ while in the second column our refined algorithm with $\varphi = 0$ is used. The results denote that the refined algorithm can find more solutions and has less iteration. In the third column, p_b is used to reduce the effect of local best positions. The results indicate that this improvement can enhance both the success rate and the reduction of iteration except for the last one (comparing with PS-CSP$_{\varphi=0}$). The hybrid algorithm HPS-CSP has the best performance on these instances. It can reduces the iteration dramatically. Both PS-CSP$_{\varphi=0,p_b}$ and HPS-CSP can find all solutions of each run.

Further random constraint satisfaction problem theories are proved there exists exactly critical values at which the phase transitions occur. The hardest instances to solve are concentrated in the sharp transition region and become remarkable benchmark to evaluate the efficiency of the newly developed algorithms. These efforts result in the Model RB [13], the state of the art in random CSP Model theory, which is a revision to the standard Model B. Those studies also revise the drawback of Model E and produce in a simple model to generate hard satisfiable instances[14][2]. Now we use the benchmark problems generated from Model RB to test our algorithm[3]. In our experiments the repetitive constraint instances are omitted. The parameter setting unchange with the above setup. The PS-CSP is completely defeated by these benchmarks and does not appear in the experimental results any longer.

[2] http://www.cril.univ-artois.fr/∼lecoutre/research/tools/RBGenerator.jar.

[3] http://www.nlsde.buaa.edu.cn/∼kexu/benchmarks/benchmarks.htm. or http://www.cril.univ-artois.fr/∼lecoutre/research/benchmarks/benchmarks.html# instances. the former also includes the solved results comparison with CSP Solver Competition.

Table 2. Comparison results with the success rate and the mean iteration of partial instances generated by Model RB

instances	ERA		SA		HPS-CSP	
	succ.	mean iter.	succ.	mean iter.	succ.	mean iter.
frb30-15.csp	26/50	166974.6	50/50	12150.8	50/50	**4390.7**
frb35-17.csp	7/50	526292.7	49/50	31310.1	**50**/50	59992.3
frb40-19.csp	12/50	343972.8	50/50	47494.1	50/50	**38372.6**
frb45-21.csp	0/50	–	35/50	462957	**50/50**	**210792**
frb50-23.csp	0/50	–	0/50	–	**50/50**	**475337**
frb53-24.csp	0/50	–	0/50	–	**41/50**	**479826**

Table 2 reports the results of the HPS-CSP algorithm on the hardest benchmark problems. For success rate, the HPS-CSP has higher success rate than ERA and SA. Our HPS-CSP can solve 30 instances out of all the 35 instances. The first 25 instances can always be tackled by HPS-CSP although on the sixth group instance there are 9 run unsolved left by HPS-CSP. For mean iteration, HPS-CSP has less iteration than ERA for these instances and, in most situations, HPS-CSP has less iteration than SA except for frb35-17 series instances.

7 Conclusion

In this paper, a more direct update formula of position is induced in our analysis, and we find φ_1 and φ_2 being set to 0 can be viewed as a heuristic like min-conflict heuristic. Our opposite experimental results seem to prove that our observation is correct. We believe that reducing the effect of local best position can improve the performance the algorithm. To improve the performance of particle swarm algorithm for CSPs, simulated annealing algorithm is introduced in the discrete particle swarm algorithm. After the update of positions, SA is activated on global best particle to get better solution quality. Also ERA model is used to generate neighbor configurations for the consideration of efficiency. Meanwhile diversity process and tabu list are combined with particle swarm algorithm for avoiding sticking in the local optima and in the useless loop throughout the search procedure. Experimental results indicate our hybrid algorithm has advantages either on search capability or on iterations.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under grant no. 60273080 and no. 60473003. This work also was supported in part by the Outstanding Youth Foundation of Jilin province of China under grant no. 20030107.

References

1. Kennedy, J, Eberhart, R. C: Particle swarm optimization. In: IEEE Int. Conf. on Neural Networks, Perth, Australia. (1995) 1942–1948
2. K.E. Parsopoulos and M.N. Vrahatis: Particle swarm optimization method for constrained optimization problems. In P. Sincak, J. Vascak, V. Kvasnicka, and J. Pospichal, editors, Intelligent Technologies—-Theory and Application: New Trends in Intelligent Technologies, volume 76 of Frontiers in Artificial Intelligence and Applications, IOS Press. (2002) 214–220
3. Hu, X. and Eberhart, R. C: Solving constrained nonlinear optimization problems with particle swarm optimization. In: Proc. of the Sixth World Multiconference on Systemics, Cybernetics and Informatics 2002 (SCI 2002), Orlando, USA. (2002) 203–206
4. Laskari, E. C., Parsopoulos, K. E., and Vrahatis, M. N: Particle swarm optimization for integer programming. In: Proc. of the IEEE Congress on Evolutionary Computation (CEC 2002), Honolulu, Hawaii, USA. (2002) 1582–1587
5. L. Schoofs, B. Naudts: Swarm intelligence on the binary constraint satisfaction problem. In: Proc. of the IEEE Congress on Evolutionary Computation (CEC 2002), Honolulu, Hawaii, USA. (2002) 1444–1449
6. M.Clerc: Discrete Particle Swarm Optimization: A Fuzzy Combinatorial Black Box. In. http://clerc.maurice.free.fr/pso/Fuzzy_Discrete_PSO/Fuzzy_DPSO.htm
7. C. Solon: Ants Can Solve Constraint Satisfaction Problems. In: IEEE Trans. Evolutionary Computation 2002, VOL. 6(4), (2002) 347–357
8. J. Kennedy and R. C. Eberhart: A discrete binary version of the particle swarm algorithm. In: Proc. of International Conference on Systems. Man, and Cybernetics, Piscataway, NJ, USA, (1997) 4104–4109
9. S.Kirkpatrick, C.D.Gelatt and M.P.Vecchi: Optimization by simulated annealing. Science, Vol. 220(13), (1983) 671–680
10. Jiming Liu, Han Jing, Y.Y.Tang: Multi-agent oriented constraint satisfaction. Artificial Intelligence, Vol. 136, (2002) 101–104
11. S.Minton, M.D.Johnston, A.B.Philips,P.Laird: Minimizing conflicts: A heuristic repair method for constraint–satisfaction and scheduling problems. Artificial Intelligence, Vol. 58, (1992) 161–205
12. D.Achlioptas, L.M.Kirousis, E.Kranakis, D.Krizanc, M.S.O.Molloy and Y.C.Stamaiou: Random constraint satisfaction: a more accurate picture. In: Proc. of CP'97, Schloss Hagenberg, Austria, (1997) 107–120
13. K. Xu and W. Li: Exact Phase Transitions in Random Constraint Satisfaction Problems. Journal of Artificial Intelligence Research, Vol. 12, (2000) 93–103
14. K. Xu, F. Boussemart, F. Hemery and C. Lecoutre: A Simple Model to Generate Hard Satisfiable Instances. In: Proc. of 19th International Joint Conference on Artificial Intelligence (IJCAI), Edinburgh, Scotland, (2005) 337–342

Global Numerical Optimization Based on Small-World Networks

Xiaohua Wang[1], Xinyan Yang[1,2], and Tiantian Su[1]

[1] Institute of Intelligent Information Processing, Xidian University, Xi'an 700071, China
[2] School of Electronics and Information Engineering, Soochow University, Suzhou 215021, China
wangxh@mail.xidian.edu.cn

Abstract. Inspired by the searching model proposed by Kleinberg in a small-world network and based on a novel proposed description that an optimization can be described as a process where information transmitted from a candidate solution to the optimal solution in solution space of problems, where the solution space can also be regarded as a small-world network and each solution as a node in the small-world network, a new optimization strategy with small-world effects was formulated in this paper. The analysis and the simulation experiments in the global numerical optimization problems indicated that the method achieved a fast convergence rate and obtained a good searching performance in optimization.

1 Introduction

The notion of small-world was first introduced in a sociological experiment to show that any pair of individuals in large social networks could be connected through a short chain of acquaintances by Stanley Milgram in 1967 [1]. As a suitable model for abstracting real-world complex networks, small-world networks (SWNs) have been demonstrated in many systems comprised by individuals with some relations. Now the research of SWNs has reached a great achievement in many fields [2]-[7].

Many search algorithms have been developed and adopted to optimization problems [8]-[10]. These algorithms have approved that they could search the global optimum with a high probability. Now emergence of SWNs supplies a new idea for optimization. For optimization searching in SWNs, Kleinberg presented a searching model in a SWN [7]. He illustrated that it was possible to use local information with limited global knowledge to find short paths between nodes without an entire network map. And all those indicated that there existed a certain speedy mode to transmit information in a SWN.

Considering an optimization, it can be described by a process where information transmitted from a candidate solution to the optimal solution in solution space of the problem. And if regarding the solution space of optimization as a SWN and each solution as a node in the SWN, a certain relationship between optimization and SWN can be obtained.

In this paper, the authors proposed a novel algorithm with small-world effects for optimization, which was named *Small-World Optimization Algorithm (SWOA)*. The

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 194 – 203, 2006.

idea behind this method was to construct a set of searching arithmetic operators based on SWNs. The theory analysis and the simulation experiments in the optimization problems indicated that the proposed method can achieve a fast convergence rate and obtain a good performance in optimization.

The rest of this paper is organized as follow. Section 2 describes how to search in small-world networks. Section 3 describes the basic SWOA. Section 4 carries out some comparisons with the other algorithms on 10 high-dimensional functions, and also discusses the experimental results. And finally Section 5 states the conclusions.

2 Searching in Small-World Networks

2.1 Modeling a Small-World Network

Usually SWNs are characterized by two distinct properties: small characteristic path length and large clustering coefficient. The characteristic path length is defined to be the average number of edges in the shortest path between a pair of nodes in the SWNs. It measures the efficiency of information propagation between nodes. The clustering coefficient measures the degree of a typical neighborhood and is defined to be the probability that two nodes connected to a common node are also connected to each other.

The research pointed out that the SWNs exhibited connectivity characteristics between random and regular lattice [7], Watts and Strogatz proposed a rewired model by randomly rewiring some edges of an ordered structure [11]. The edges in the underlying ordered structure can be viewed as local short contacts. By rewiring edges, it not only increases the randomness of the structure but also creates a few long-range contacts. Thus with a high clustering subnet comprising of the local short contacts and a few random long-range contacts, the small-world networks can be established.

Another model was proposed by Kleinberg [7] and illustrated by an $n \times n$ lattice with some non-negative constants δ, q and r. Denote the distance between nodes u and v as $D(u,v)$. The edges can be constructed as follows: the node u has a local short contact to nodes v if $D(u,v) \le \delta$, and q other edges, the long-range contacts, to other nodes w are on the probability proportional to $D(u,w)^{-r}$. The parameter r partially reflects the offset degree of a small-world lattice to a random or regular lattice.

2.2 Searching Mechanism in Small-World Network

Considering a SWN $G = (V, E)$, where V is the set of nodes and E is the set of edges. A node v is called a neighbor of node u if there is an edge between them, and $v, u \in V$. All the neighbors of u comprise the neighborhood of u. The location of node u is the coordinate that shows the position of u. Since a SWN is comprised by individuals with some relations, the node u can be characterized by an individual feature ξ_u. And the feature distance, $D(\xi_u, \xi_v)$, measures the difference between u and v in terms of the node individual features.

Given a source node S with individual features ξ_S and a destination D with individual features ξ_D in a SWN, a search is to determine whether there is a path from S to a given node with desired individual feature ξ_D or not. Assuming that a node with individual features ξ_D always exists, the problem can be simplified to find a mechanism to reach the destination. The Kleinberg's search mechanism is based on local optimality and can find an approximate shortest path to a destination. In searching, there is no any global and centralized mechanism to control the search direction. For example, when the search is passed to a node u, the next node to reach is determined by the local information of u. It was possible to use local information with limited global knowledge to find short paths between nodes, and it was a feasible choice in searching a large network without an entire network map [7].

In order to perform a search in a SWN, the authors first presume that each node in the network possesses two kinds of information: the global and the local information. The global information includes the destination feature, and the method of calculating the feature distance, while the local information includes the individual feature and locations of its own, and individual feature of its neighbor nodes.

3 Small-World Optimization Algorithm (SWOA)

The two models above were the basic models for simulating the real-world networks with small-world characteristics. The correlation between the local short contact and long-range contacts provided an important clue to transmit information to a given destination in a SWN effectively, namely information transmission in the SWN would become high-efficiency when a few random long-range contacts were added into the local short contacts [7] [12]. Therewith, how to construct a local search arithmetic operator based on local short contacts and a global search arithmetic operator based on a few random long-range contacts became very important. Here the authors added the global search, which was resulted from the random long-range contacts mechanism, on the base of the local search, which was resulted from the local short contacts mechanism. Then the optimization strategy based on small-world networks was formulated.

3.1 Useful Definitions of SWOA

In general, a global numerical optimization could be formulated as solving the following objective function.

$$\text{minimize } f(\boldsymbol{X}),\ \boldsymbol{X} = (x_1, \cdots, x_m) \in \boldsymbol{I},\ \boldsymbol{d} \le X \le \boldsymbol{u}, \tag{1}$$

where $\boldsymbol{I} \subseteq R^m$ defined the search space which was an m-dimensional space bounded by the parametric constraints $x_i \in [d_i, u_i]\ (i = 1, 2, \cdots, m)$. Thus $\boldsymbol{I} = [\boldsymbol{d}, \boldsymbol{u}]$, where $\boldsymbol{d} = (d_1, \cdots, d_m)$ and $\boldsymbol{u} = (u_1, \cdots, u_m)$.

According to Milgram's sociological experiment, in which different individuals was chosen to be source nodes who would be given a letter to deliver to a given target person [1], a candidate solution aggregate $\boldsymbol{S}$ is adopted as a set of source nodes instead of picking a single node in solution space which has been regarded as a SWN. Namely $\boldsymbol{S} = \{S_1, S_2, \ldots, S_n\}$, where n is the number of nodes in $\boldsymbol{S}$. Using some nomenclature in evolutionary computation for reference, the authors introduce the similar nomenclature here. Thus $\boldsymbol{S}$ is the search population, subscript n is the search population size, and S_i is the i^{th} individual in $\boldsymbol{S}$, and denoted the code of variant $\boldsymbol{X}$ of the objective function. Given an m-dimensional optimization problem, each node S_i would be divided into m segments, and each segment indicated x_i respectively. The authors write $S_i = (x_{i1}, x_{i2}, \ldots, x_{im})$ here.

Define the optimal destinations solution space of the objective function as:

$$\boldsymbol{O} \triangleq \{S_i \in \boldsymbol{I} \mid f(S_i) = f^* \equiv \min(f(\boldsymbol{X}) \mid \boldsymbol{d} \le \boldsymbol{X} \le \boldsymbol{u}), 1 \le i \le n\}, \tag{2}$$

where f^* denotes the optimal solution of the objective function.

Define the individual features distance between S_i and S_j as $D(S_i, S_j) \triangleq \left\| S_i - S_j \right\|$ in terms of the coding form, for instance, Hamming distance is adopted for the binary code, and Euclidean distance for decimal code.

Define the δ neighborhood of S_i as

$$\mu^{\delta}(S_i) \triangleq \{S_j \mid 0 < D(S_i, S_j) \le \delta\}, \tag{3}$$

where δ denotes the threshold value of neighborhood. $|\mu^{\delta}(S_i)|$ denotes the number of members in the $\mu^{\delta}(S_i)$.

Define the δ non-neighborhood of S_i as

$$\overline{\mu^{\delta}(S_i)} \triangleq \{S_j \mid D(S_i, S_j) > \delta\}, \tag{4}$$

With the useful definitions above, a mapping relationship between optimization and search in a SWN can be found. The solution space of the optimization can be regarded as a SWN and each solution as a node in the SWN. And in term of the δ value the authors can distinguish between the local short contacts and long-range contacts in the SWN.

3.2 Local Short Contacts Search Arithmetic Operator

This arithmetic operator was defined as Ψ, and its main function was to transmit information from $S_i(k)$ to $S_i(k+1)$ which was the closest node to $\boldsymbol{O}$ in the $\mu^{\delta}(S_i(k))$, the authors write $S_i(k+1) \leftarrow \Psi(S_i(k))$. And the arithmetic operator Ψ was described as the followings:

Step1. Set δ, and get $n_\delta = rand \times |\mu^\delta(S_i(k))|$, where *rand* denotes a random number in the domain [0,1], obviously $n_\delta \le |\mu^\delta(S_i(k))|$.

Step2. $n_\delta^{'} = 0$, and $S_i^{'}(k) \leftarrow S_i(k)$ and $S_i^{''}(k) \leftarrow S_i(k)$.

Step3. $S_i^{'}(k) \leftarrow S_i^{''}(k)$, and execute **operating** $\boldsymbol{M}_{AEA}(S_i^{'}(k))$;

Step4. For the objective function, If $f(S_i^{'}(k)) < f(S_i(k))$, execute $S_i(k) \leftarrow S_i^{'}(k)$;

Step5. $n_\delta^{'} = n_\delta^{'} + 1$; If $n_\delta^{'} = n_\delta$, execute $S_i(k+1) \leftarrow S_i(k)$, and the iteration is terminated, otherwise, return to **Step3**.

Operating $\boldsymbol{M}_{AEA}(S_i)$ is similar to AEA mutation [13], and the operation method is implemented as follows**:**

$$\boldsymbol{M}_{AEA}(S_i) = (x_{i1}, \cdots, \boldsymbol{M}_{AEA}(x_{ij}), \ldots, x_{im}), \tag{5}$$

$$\boldsymbol{M}_{AEA}(x_{ij}) = x_{ij} - (fix(x_{ij} \times 10^{rnd(15)}) \bmod 10 - rnd(10)) \Big/ 10^{rnd(15)}, \tag{6}$$

where x_{ij} is chosen in S_i randomly, $fix(x)$ denotes taking integer part of x, and $rnd(N)$ denotes getting a random integer less than N.

3.3 Random Long-Range Connections Search Arithmetic Operator

This arithmetic operator is defined as Γ, and its main function is to select a random node in $\overline{\mu^\delta(S_i)}$ as the next destination node $S_i^{'}(k)$, obviously $D(S_i(k), S_i^{'}(k)) > \delta$. The authors write $S_i^{'}(k) \leftarrow \Gamma(S_i(k))$, here $S_i(k)$ is not the global optimal node for the objective function $f(S_i(k))$ at present. And the arithmetic operator Γ is described as the followings:

$$S_i^{'}(k) \leftarrow \Gamma(S_i(k)) = \begin{cases} \boldsymbol{I}_R(S_i(k)) & \text{if } rand \le 0.5 \\ \boldsymbol{I}_B(S_i(k)) & \text{otherwise} \end{cases}. \tag{7}$$

Operating $\boldsymbol{I}_R(S_i(k))$ is implemented as follows**:**

Under the two point p and q (suppose $p<q$) randomly chosen in $S_i(k) = \{x_{i1}, x_{i2}, \cdots, x_{ip}, \cdots, x_{iq}, \cdots, x_{im}\}$, the offspring of $S_i^{'}(k)$ will be $S_i^{'}(k) = \{x_{i1}, x_{i2}, \cdots x_{iq}, x_{iq\text{-}1} \cdots, x_{ip+1}, x_{ip}, \cdots, x_{im}\}$.

Operating $\boldsymbol{I}_B(S_i(k))$ is implemented as follows**:**

Execute binary coding operation of the $S_i(k)$, and get a binary cluster $A_i = \{a_1, \cdots a_l\}$ with the length l. Taking the required precision in SWOA into account here, l is set to $20 \times m$. And the coding method is given by:

$$a_i = (x_{ij} - d_i) \times (2^{l_c} - 1) \big/ (u_i - d_i). \tag{8}$$

Under the random integers p and q (suppose $1 \le p < q \le \delta$, and $|p-q| > \delta$) randomly chosen in $A_i = \{a_1, \cdots, a_p, \cdots a_q, \cdots a_l\}$, the offspring of $A_i^{'}$ will be $A_i = \{a_1, \cdots, a_q, a_{q-1}, \cdots a_{p+1}, a_p, \cdots a_l\}$. Then with the decoding method implemented as formula (9), the authors can get $S_i^{'}(k) = \{x_{i1}^{'}, x_{i2}^{'}, \cdots, x_{i3}^{'}\}$.

$$x_{ij}^{'} = a_j^{'} \times (u_i - d_i) \big/ (2^{l_c} - 1) + d_i . \tag{9}$$

3.4 Small-World Optimization Algorithm (SWOA)

Based on the expatiation above, the authors introduced SWOA shown as the followings:

Step1. Initialize: Set population size *n*, probability P_{short} for the local short connections search, probability P_{os} for optimization of search population, and the maximal generation threshold *mgt*. Then generate an original search population $\boldsymbol{A}(0) = \{\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_n\}$ with the size *n* in the domain [0,1] randomly, where $\boldsymbol{A}_i = \{a_{i1}, a_{i2}, \ldots, a_{im}\}$.

Step2. Domain transformation: The method is implemented as follows.

$$x_{ij} = (u_i - d_i) \times a_{ij} + d_i,\ x_{ij} \in S_i,\ i = 1, 2 \cdots n,\ j = 1, 2 \cdots m . \tag{10}$$

Then $\boldsymbol{S}(0) = \{S_1(0), S_2(0), \ldots, S_n(0)\}$ can be obtained.

Step3. Optimize the search population: The method is implemented as follows.

If $rand \le P_{os}$, execute $S^{'} = (x_1^*, x_2^*, \cdots x_m^*)$, and

$$x_j^* = \{x_{ij} \mid f(x_{11}, x_{12}, \cdots x_{ij}, \cdots x_{1m}) = Min^{'},\ i = 0, 1 \cdots n\} . \tag{11}$$

where $Min^{'}$ denotes the objective function get the local optimal value for the different $x_{ij}, i = 0, 1 \cdots n$, then the worst node of the current generation would be replaced by $S^{'}$; otherwise continue.

Step4. Calculate the value of objective function: The value of objective function is defined as follow:

$$F(k) : \{f(\boldsymbol{S}(k))\} = \{f(S_1(k)), f(S_2(k)), \cdots, f(S_n(k))\} . \tag{12}$$

Step5. Check the halted condition of algorithm: If the condition is met, then the algorithm terminates, otherwise continue.

Step6. Hold the optimal node: If the optimal node of the last generation is better than that of the current generation, the worst node of the current generation would be replaced by that of the last generation.

Step7. $i \leftarrow 0$, If $rand \le P_{short}$, execute $S_i(k+1) \leftarrow \Psi(S_i(k))$ and check the halted condition of algorithm; otherwise execute $S_i^{'}(k) \leftarrow \Gamma(S_i(k))$. If the condition is met here, the SWOA algorithm terminates.

Step8. $i = i+1$. If $i < n$, return to **Step7**, otherwise continue.
Step9. $k = k+1$. If $k < mgt$, return to **Step3**, otherwise the SWOA algorithm terminates.

In this paper, the halted conditions of SWOA are defined as follows:

- The generation times have reached the maximal generation threshold;
- $\left| f^* - f^{best} \right| < \varepsilon$, where f^* is the global optima of f, f^{best} is the current optima, and ε is the required precision.

4 Experimental Studies

4.1 Test Functions

In order to prove the effect of the novel algorithms proposed, the authors use ten benchmark functions in the experiments shown as follows.

$$f_1(x) = \sum_{i=1}^{N} x_i^2 \quad -100 \le x_i \le 100 \tag{13}$$

$$f_2(x) = \sum_{i=1}^{N} |x_i| + \prod_{i=1}^{N} |x_i| \quad -10 \le x_i \le 10 \tag{14}$$

$$f_3(x) = \sum_{i=1}^{N} (\sum_{j=1}^{i} x_j)^2 \quad -100 \le x_i \le 100 \tag{15}$$

$$f_4(x) = \sum_{i=1}^{N} (\lfloor x_i + 0.5 \rfloor)^2 \quad -100 \le x_i \le 100 \tag{16}$$

$$f_5(x) = \sum_{i=1}^{N} i x_i^4 + random[0,1) \quad -1.28 \le x_i \le 1.28 \tag{17}$$

$$f_6(x) = \frac{1}{N} \sum_{i=1}^{N} (x_i^4 - 16x_i^2 + 5x_i) \quad -5 \le x_i \le 5 \tag{18}$$

$$f_7(x) = \sum_{i=1}^{N} (x_i^2 - 10\cos(2\pi x_i) + 10) \quad -5.12 \le x_i \le 5.12 \tag{19}$$

$$f_8(x) = -\sum_{i=1}^{N} x_i \sin(\sqrt{|x_i|}) \quad -500 \le x_i \le 500 \tag{20}$$

$$f_9(x) = \sum_{i=1}^{N} \frac{x_i^2}{4000} - \prod_{i=1}^{N} \cos(\frac{x_i}{\sqrt{i}}) + 1 \quad -600 \le x_i \le 600 \tag{21}$$

$$f_{10} = -20\exp\left(-0.2\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}\right) - \exp\left(\frac{1}{N}\sum_{i=1}^{N} \cos(2\pi x_i)\right) + 20 + e \quad -30 \le x_i \le 30 \tag{22}$$

f_1-f_5 are unimodal functions, and f_6-f_{10} are multimodal functions where the number of local minima increases with the problem dimension. For example, the number of local minima of f_7 is about $10N$ in the given search space.

4.2 Experimental Setup

In order to prove the effect and efficiency of our algorithm, the authors present some simulation results and comparisons. The performance of the SWOA is evaluated via computer simulations and compared with that of APPA [14], OGA/Q[15]. In the experiments, the parameters of SWOA are set as: $n=15$, $\delta=1$, $P_{short}=0.8$, $P_{os}=0.4$, and $mgt=10000$.

4.3 Comparison Between SWOA, APPA and OGA/Q on Functions

The authors regard the mean number of function evaluations and the mean function values as objective. The statistical results of the different algorithms over 20 trials respectively are summarized in Table 1. Where dimension m=30, except for f_6 with 100.

It is indicated from Table 1 that for all test functions, SWOA can find the global optimal solution with the required precision over 20 trials. As also can be seen, the mean number of evaluations of SWOA is lesser than that of APPA and OGA/Q except for function f_3, f_4, f_9. And as a result of the halted conditions of SWOA, the authors can see the mean functions values of SWOA may be worse than APPA and OGA/Q for some functions, but considering the computational cost shown by the mean number of evaluations, SWOA is competent for the numerical optimization problems.

Table 1. Comparison between SWOA, APPA and OGA/Q on functions with 30 dimensions except for f_6 with 100. "Prob" is the contrastive problems, "m" denotes the dimension of the function, "f_{min}" is the global optimal value of the function, and " " is the required precision of the function.

	Prob			Mean number of function evaluations			mean function values (mean standard deviation of function values)		
f	m	f_{min}	ε	SWOA	APPA	OGA/Q	SWOA	APPA	OGA/Q
f_1	30	0	10^{-4}	**5096.6**	5,117	112,599	3.5208×10^{-5} (3.5992×10^{-5})	2.809×10^{-7} (2.969×10^{-7})	0 (0)
f_2	30	0	10^{-4}	**3496.6**	4,220	112,612	1.8775×10^{-5} (2.8555×10^{-5})	3.579×10^{-9} (3.362×10^{-9})	0 (0)
f_3	30	0	10^{-4}	149312	**47,406**	112,576	9.9381×10^{-6} (7.8718×10^{-6})	3.826×10^{-10} (4.340×10^{-10})	0 (0)
f_4	30	0	10^{-4}	2292.1	**1,534**	62,687	0 (0)	0 (0)	0 (0)
f_5	30	0	10^{-4}	**1077.8**	24,261	112,652	4.3201×10^{-5} (2.8312×10^{-5})	4.848×10^{-4} (3.298×10^{-4})	6.301×10^{-3} (4.069×10^{-4})
f_6	100	-78.3323	10^{-3}	**6302.6**	11,541	245,930	-78.332 (1.8382×10^{-5})	-78.33009 (8.381×10^{-4})	-78.33000296 (6.288×10^{-3})
f_7	30	0	10^{-2}	**4296.2**	4,657	224,710	4.2009×10^{-3} (3.2452×10^{-3})	1.10645×10^{-10} (1.359×10^{-10})	0 (0)
f_8	30	-12569.5	10^{-2}	**4971.8**	9,499	302,166	-12569.4932 (5.3184×10^{-4})	-12569.49 (1.253×10^{-5})	-12569.4537 (6.447×10^{-4})
f_9	30	0	10^{-4}	30618	**8,280**	134,000	4.0518×10^{-5} (3.4054×10^{-5})	5.12923×10^{-15} (3.463×10^{-15})	0 (0)
f_{10}	30	0	10^{-3}	**4224.6**	8970	112421	3.0287×10^{-4} (3.8906×10^{-4})	1.95399×10^{-15} (2.398×10^{-15})	4.440×10^{-16} (3.989×10^{-17})

4.4 Discussions

According to the experimental results, the SWOA is more robust and efficient than APPA and OGA/Q for most problems the authors tested. The local short contacts search arithmetic operator and the random long-range contacts search arithmetic operator in SWOA can be regarded as the local search arithmetic operator and the global search arithmetic operator respectively. SWOA not only can find the global optima with the "collected" local optima by the local search around the neighborhood of an individual but also can jump out of the local optima by the global search, and the combined effects of the local search and the global search are capable to improve the convergence rate, and also avoid the premature convergence availably. Furthermore, the randomicity in SWOA can keep the diversities in search populations as possible.

5 Conclusions

Inspired by the SWN model proposed by Kleinberg and due to the novel proposed description of the optimization problems, a novel numerical optimization algorithm, SWOA, has been proposed in this paper. Referring to the searching mechanism in SWNs, the authors construct the local short contacts search arithmetic operator for the local short contacts search and the random long-range contacts search arithmetic operator for the global search correspondingly. The comparisons with APPA and OGA/Q considering the mean number of function evaluations and mean function values show that SWOA can provide more efficient performance with better reliability on most high-dimensional multimodal functions in test.

To summarize, SWOA can converge to the global optimal solution successfully, and still has a faster convergence rate, a comparatively better capability in optimization, namely SWOA obtains a good performance. The successful application in the global numerical optimization problems with SWOA shows that this algorithm is feasible to solve the complex problems.

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China under Grant 60133010, 60472084 and 60372045, and by the Major State Basic Research Development Program of China under Grant 2001CB309403.

References

1. S.Milgram, The Small-World Problem, Psychology Today 1, 1967, 60-67.
2. Watts, D. J. and Strogatz, S. H. Collective dynamics of small-world networks. Nature, 1998, 393: 440-442.
3. Sen P, Dasgupta P, Chatterjee A, et al. Small-world properties of the Indian railway network. Phys Rev E, 2003, 67, 036106.
4. Moore C, Newman M E J. Epidemics and percolation in small-world networks. Phys Rev E, 2000, 61, 5678-5682.

5. Newman M E J, Watts D J. Scaling and percolation in the small-world network model. Phys Rev E, 1999, 60, 7332-7342.
6. M. E. J. Newman and D. J. Watts, “Renormalization group analysis of the small-world network model”, Phys. Lett. A, vol. 263, pp. 341-346, 1999.
7. Jon Kleinberg. The Small-World Phenomenon and Decentralized Search [J]. SIAM New, 2004, 37(3):1-1.
8. Jing Liu, Weicai Zhong, and Licheng Jiao. A Multiagent Evolutionary Algorithm for Constraint Satisfaction Problems. IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 36, NO. 1, FEBRUARY 2006.
9. Licheng Jiao, Jing Liu, and Weicai Zhong. An Organizational Coevolutionary Algorithm for Classification. IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, VOL. 10, NO. 1, FEBRUARY 2006.
10. Licheng Jiao, and Lei Wang. A Novel Genetic Algorithm Based on Immunity. IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, VOL. 30, NO. 5, SEPTEMBER 2000.
11. Watts D. J., Small worlds. Princetown University Press, 1999.
12. Kleinberg J., Navigation in a small world. Nature, 406:845, 2000.
13. Mühlenbein H., Schlierkamp V. D. Predictive models for the breeder genetic algorithm. Evol. Computat., vol. 1, No.1 (1993) 25 – 49.
14. Du H. F., et al. Adaptive Polyclonal Programming Algorithm with application. ICCIMA (2003) 350 – 355.
15. Leung Y.W., Wang, Y. P. An orthogonal genetic algorithm with quantization for global numerical optimization. IEEE Trans. Evol. Comput., vol.5, No.2 (2001) 41 -53.

Real-Time Global Optimal Path Planning of Mobile Robots Based on Modified Ant System Algorithm

Guanzheng Tan and Dioubate Mamady I

School of Information Science and Engineering, Central South University
Changsha 410083, Hunan Province, China
tgz_csu@yahoo.com.cn, dioubate01@yahoo.com.cn

Abstract. A novel method for the real-time global optimal path planning of mobile robots is proposed based on the modified ant system (AS) algorithm. This method includes three steps: the first step is adopting the MAKLINK graph theory to establish the free space model of the mobile robot, the second step is adopting the Dijkstra algorithm to find a sub-optimal collision-free path, and the third step is adopting the modified AS algorithm to adjust and optimize the location of the sub-optimal path so as to generate the global optimal path. The results of simulation experiments confirm that the proposed method is effective and has better performance in convergence speed, solution variation, dynamic convergence behavior, and computation efficiency as compared with the path planning method based on the real-coded genetic algorithm.

1 Introduction

The global optimal path planning is a key problem in the navigation of autonomous mobile robots, which means finding the optimal collision-free path between a starting point and a goal in a given environment according to some criterion. At present, there are many solving methods about this problem, such as the potential field method [1] and grid method [2]. With the rapid development of evolutionary computation [3], [4], many intelligent algorithms were applied to this problem, such as neural network [5], genetic algorithms [6], the fuzzy logic and reinforcement learning [7], and so on.

Ant system (AS) algorithm is a novel kind of simulated evolutionary algorithm [8]. Its main characteristics include positive feedback search mechanism, distributed computation, and the use of a constructive greedy heuristic. So far, AS algorithm has been used successfully to solve many practical optimization problems.

This paper presents a modified AS algorithm, which is used to solve the problem of the real-time global optimal path planning of mobile robots. The global path planning includes two sub-problems: the free space modeling and path finding. In this paper, first we adopt the MAKLINK graph theory to establish the free space model of a mobile robot, then we use the Dijkstra algorithm to find a sub-optimal collision-free path, finally we use the modified AS algorithm to adjust and optimize the location of the sub-optimal path so as to obtain the global optimal path.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 204 – 214, 2006.

2 Free Space Modeling of Mobile Robot

The MAKLINK graph theory [9] is adopted to establish the free space model of a mobile robot. Fig.1 illustrates a moving environment of a mobile robot. Between the starting point *S* and goal *T*, an area with 300×300 square meters is enclosed, in which five obstacles are distributed. The boundaries of every obstacle are expanded by an amount that is equal to half of the greater size in the length and width of the robot's body plus the minimum measuring distance of the used sensors. In this case, the robot can be regarded as a "spot" with no size. In Fig.1, for each of the obstacles, the black polygon denotes its original size, and the white margins denote its expanded parts. A black part plus its white margins constitutes a so-called "grown obstacle".

In Fig.1, the symbols B_1, B_2, …, and B_{20} denote respectively the vertices of these grown obstacles. The (x, y) coordinates of B_1, B_2, …, and B_{20} are (40, 288), (66, 288), (66, 151), (40, 151), (115, 275), (95, 214), (123, 163), (170, 245), (90, 106), (90, 45), (183, 45), (183, 106), (238, 311), (212, 248), (274, 268), (258, 205), (234, 190), (234, 111), (296, 111), and (296, 137), respectively. The (x, y) coordinates of the starting point *S* and the goal *T* are (15, 335) and (315, 35), respectively.

The free space of a mobile robot consists of some polygonal areas, each of which is enclosed by several free MAKLINK lines. A free MAKLINK line is defined as: (1) whether its two end points are two vertices on two different grown obstacles or one of the two points is a vertex of a grown obstacle and the other is located on a boundary of the environment; (2) every free MAKLINK line cannot intersect any one of the grown obstacles. If we draw all the free MAKLINK lines on the environment map, we can obtain the so-called MAKLINK graph of the environment. Fig.1 also shows the MAKLINK graph of this example.

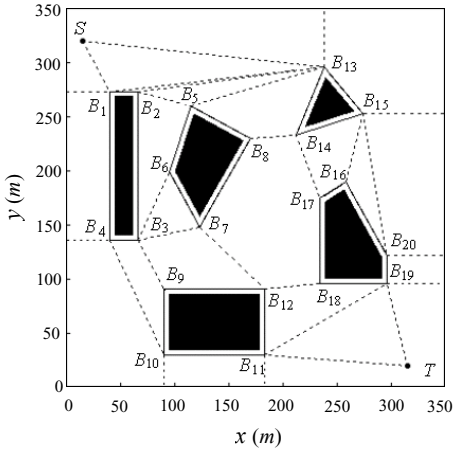


Fig. 1. An environment of a mobile robot and its MAKLINK graph

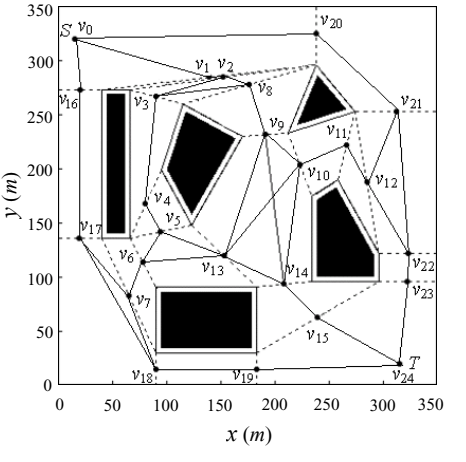


Fig. 2. Network graph for free motion of the mobile robot

As shown in Fig.2, assume that the middle points of all free MAKLINK lines on a MAKLINK graph are denoted respectively by v_1, v_2, …, v_l (where l=23). If each pair of middle points on two adjacent free MAKLINK lines are connected together, a network graph can be formed, which gives the possible paths for the free motion of the mobile robot. In Fig.2, points S and T are also denoted by v_0 and v_{l+1}, respectively. In fact, Fig.2 is an undirected graph. We use G (V, E) to denote this graph, where V=$\{v_0, v_1, \ldots, v_{l+1}\}$; E is a set of lines which includes: the lines, each of which connects a pair of the middle points on two adjacent free MAKLINK lines; the lines connecting S and the middle points on the free MAKLINK lines adjacent to S; and the lines connecting T and the middle points on the free MAKLINK lines adjacent to T. In this paper, the undirected graph G (V, E) is used as the free space model.

3 Searching for Sub-optimal Path Using Dijkstra Algorithm

The Dijkstra algorithm is widely used to search for the shortest path on a network graph. When using it on the network graph G (V, E), the path generated by it is passing through the middle points of the relevant free MAKLINK lines. So, the generated path is only a sub-optimal path.

Before using this algorithm, it is necessary to define the adjacency matrix with weights for the network graph G (V, E). Each element of the matrix represents the length of the straight line between two adjacent path nodes on G (V, E), where a path node means the intersection of a robot's moving path and a free MAKLINK line. For the problem discussed here, each element of the adjacency matrix is defined as:

$$adjlist[i][j] = \begin{cases} length\,(v_i, v_j), & \text{if } edge\,(v_i, v_j) \in E \\ \infty, & \text{others} \end{cases}, \tag{1}$$

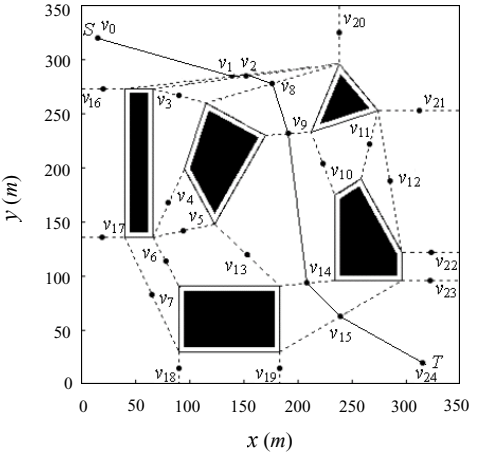


Fig. 3. Sub-optimal path generated by Dijkstra algorithm

where *adjlist* [*i*][*j*] is the element corresponding to the *i*th row and the *j*th column of the matrix. Using the Dijkstra algorithm the sub-optimal path for the example in Fig.1 is obtained as: $S \rightarrow v_1 \rightarrow v_2 \rightarrow v_8 \rightarrow v_9 \rightarrow v_{14} \rightarrow v_{15} \rightarrow T$, as shown in Fig.3. The length of this path is 485.851 meters.

4 Modified AS Algorithm Based Global Optimal Path Searching

In this section, we will adopt the modified AS algorithm to optimize the locations of the path nodes on the sub-optimal path to obtain the global optimal path.

4.1 Description of Optimization Problem

Assume that the sub-optimal path is denoted in order by nodes P_0, P_1, P_2, …, P_d, and P_{d+1}, where P_0 and P_{d+1} denote the starting point *S* and the goal *T* respectively. At the beginning, these path nodes lie on the middle points of the relevant free MAKLINK lines. Now, we need to adjust and optimize their locations on their corresponding free MAKLINK lines. The adjustment method is described as follows.

For a path node P_i, i=1, 2, …, d, its location is on the free MAKLINK line $P_{i1}P_{i2}$. Introducing a parameter h_i, the location of P_i on line $P_{i1}P_{i2}$ can be expressed as:

$$P_i = P_{i1} + (P_{i2} - P_{i1}) \times h_i, \quad h_i \in [0,\ 1], \quad i = 1,2,\cdots,d. \tag{2}$$

Obviously, given a set of values to {h_1, h_2, …, h_d}, together with the staring point *S* and the goal *T*, a robot path can be generated. Rewriting P_i as P_i (h_i), then the objective function of the optimization problem can be defined as:

$$L = \sum_{i=0}^{d} length\{P_i(h_i), P_{i+1}(h_{i+1})\}, \tag{3}$$

where *length* {P_i (h_i), P_{i+1} (h_{i+1})} represents the straight-line distance between P_i and P_{i+1}. Our purpose is to use the modified AS algorithm to find the optimal parameter set {h_1^*, h_2^*, …, h_d^*} so that the objective function *L* has the minimum value.

4.2 Generating Method of Ant Moving Paths

In order to use the AS algorithm conveniently, we express the values of parameters h_1, h_2, …, and h_d on plane O-**XY**. As shown in Fig.4, we draw *d* lines on O-**XY** which have equal length and equal interval and are perpendicular to axis **X**. The *x* coordinates of these lines are represented by 1, 2, …, and *d*. Then, we divide each of these lines equally into ten portions, and thus eleven nodes are generated on these lines. The *y* coordinates of the eleven nodes on each line are 0, 0.1, 0.2, …, and 1.0, which represent the eleven possible values of parameter h_i, i=1, 2, …, *d*. Corresponding to the MAKLINK graph, this means each of the relevant free MAKLINK lines is transformed into the eleven discrete path nodes, which can be chosen easily by an ant.

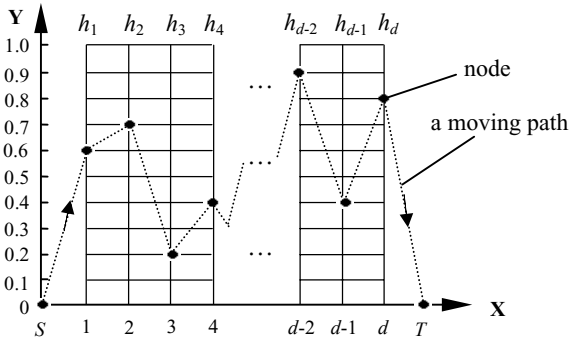


Fig. 4. Generating of nodes and moving paths

Fig.4 is a grid graph on which there are d×11 nodes. Using n_{ij} denotes node j on line h_i, i=1, 2, …, d; j =0, 1, …, 10. Let an ant depart from the starting point S. In its each step forward, it chooses a node from the next line h_i, i=1, 2, …, d, and then moves to this node along the straight line. When it arrives at the goal T, it completes one tour and its moving path can be expressed as $Path$={S, n_{1j}, n_{2j}, …, n_{dj}, T}.

Assume that from any node on line h_i to any node on the next line h_{i+1}, each ant in the ant colony has the same moving time. So, if all ants depart from the point S at the same time, they will arrive on each line h_i at the same time too, and finally arrive at the goal T at the same time. To this moment, the AS algorithm completes an iteration.

Let $\tau_{ij}(t)$ represent the concentration of pheromone at node n_{ij}, where t is the iteration counter of the AS algorithm. Assume that at initial time t=0 all the nodes have the same pheromone concentration τ_0. Assume that the total number of ants is m. In moving process, for an ant k, when it locates on line h_{i-1}, it will choose a node j from the next line h_i to move to according to the following random transition rule:

$$P_{ij}^{k} = \frac{[\tau_{ij}(t)]^{\alpha} \cdot [\eta_{ij}]^{\beta}}{\sum_{j=0}^{10} [\tau_{ij}(t)]^{\alpha} \cdot [\eta_{ij}]^{\beta}}, \tag{4}$$

where η_{ij} represents the visibility of node n_{ij} and is computed by (5); α, β are two adjustable parameters. According to the suggestion in [8], α and β are set to 1 and 5 respectively. For visibility η_{ij}, we define it as:

$$\eta_{ij} = \frac{1.1 - \left| y_{ij} - y_{ij}^{*} \right|}{1.1}, \tag{5}$$

where y_{ij} is the y coordinate of node n_{ij}; the values of y_{ij}^{*} are set in the following way. In the first iteration of the AS algorithm, the values of y_{ij}^{*} are set respectively to the values of h_1, h_2, …, and h_d that are corresponding to the nodes on the free MAKLINK lines found by the Dijkstra algorithm. In each of the following iterations, the values of y_{ij}^{*} are set to the values of h_1, h_2, …, and h_d which are corresponding to the path nodes on the optimal path generated by the AS algorithm in the previous iteration.

4.3 Updating Rules of Pheromone Concentration

For an ant k, when it passes through a node n_{ij}, the pheromone concentration $\tau_{ij}(t)$ of this node needs to be updated immediately using the *local updating rule* [10]:

$$\tau_{ij}(t) \leftarrow \rho \cdot \tau_{ij}(t) + (1-\rho) \cdot \tau_0 , \tag{6}$$

where $0<\rho<1$ is the pheromone decay coefficient. When the AS algorithm completes an iteration, all the ants arrive at goal T. Now, the pheromone concentration of each node on the grid graph needs to be updated according to the *global updating rule*:

$$\tau_{ij}(t) \leftarrow \rho \cdot \tau_{ij}(t) + (1-\rho) \cdot \Delta\tau_{ij}(t) + \gamma , \tag{7}$$

where $\Delta\tau_{ij}(t)$ is computed by

$$\Delta\tau_{ij}(t) = \begin{cases} 1/L^+ , & \text{if the shortest robot path } T^+ \text{ generated in the} \\ & \text{current iteration passes through node } n_{ij} \\ 0 , & \text{otherwise.} \end{cases} , \tag{8}$$

where L^+ is the length of the path T^+, the shortest path in the current iteration. γ in (7) is a penalty factor newly introduced by us and computed by the following formula:

$$\gamma = (L^* - L^+) \cdot | y_i(T^*) - y_{ij}(T^+) | / Q , \tag{9}$$

where L^* is the length of the path T^* which is the shortest robot path generated from the beginning of the trial till the previous iteration; $y_i(T^*)$ is the y coordinate of the node on line h_i which belongs to the path T^*; $y_{ij}(T^+)$ is the y coordinate of the node n_{ij} which belongs to the path T^+; Q is a positive constant.

The penalty factor γ is applied only at each node of the path T^+. It can guide the ants to search for the better paths than path T^*.

4.4 Modified AS Algorithm for Global Optimal Robot Path Searching

The modified AS algorithm can be summarized as follows.

Step 1: Establish the free space model using the MAKLINK graph theory, and then find a sub-optimal collision-free path using the Dijkstra algorithm.

Step 2: Define ant number m and specify the values of α, β, ρ, τ_0, and Q; For each ant k (k=1, 2, …, m), define a one-dimensional array $Path_k$ with d elements, in which the y coordinates of the d nodes that ant k will pass through on the grid graph Fig.4 will be stored in order. Array $Path_k$ can be used to denote the moving path of ant k.

Step 3: Set the iteration counter t=1 and define the maximum number of iterations NC; Then place all the m ants at the starting point S.

Step 4: Set i=1.

Step 5: Set k=1.

Step 6: Select a node on line h_i for ant k using (4) and *Roulette Wheel Selection Method*, move ant k to this node, and save the y coordinate of this node into the ith element of $Path_k$; Then, update the pheromone concentration of this node using the local updating rule (6).

Step 7: Set $k \leftarrow k+1$. If $k \le m$, go to *Step* 6; Otherwise, go to *Step* 8.

Step 8: Set $i \leftarrow i+1$. If $i \le d$, go to *Step* 5; Otherwise, go to *Step* 9.

Step 9: Move each ant from its location on line h_d to the goal T.

Step 10: For each ant k: (a) according to array $Path_k$, obtain the values of parameter set $\{h_1^k, h_2^k, \ldots, h_d^k\}$; (b) according to $\{h_1^k, h_2^k, \ldots, h_d^k\}$ determine the locations of path nodes $P_1^k, P_2^k, \ldots$, and P_d^k on their corresponding free MAKLINK lines; and (c) using (3), compute the length L_k of the robot path found by ant k.

Step 11: Compare these m paths and find the shortest robot path T^+ in the current iteration t. Compare T^+ with T^*, the shortest path generated from the beginning of the trial till the previous iteration $t-1$ (for the first iteration, directly denote its shortest path by T^*), and denote the new shortest robot path by T^*, then save the values of h parameter set corresponding to T^* into $\{h_1^*, h_2^*, \ldots, h_d^*\}$.

Step 12: Set each element of $Path_k$ to zero, k=1, 2, …, m.

Step 13: Update the pheromone concentration of each node on the grid graph Fig.4 using the global updating rule (7).

Step 14: Set $t \leftarrow t+1$. If $t < NC$ and all of the m ants do not make the same tour, place all the ants at the starting point S and go to *Step* 4; If $t < NC$ but all of the m ants make the same tour or $t=NC$, output the optimal (shortest) robot path T^* and its corresponding parameter set $\{h_1^*, h_2^*, \ldots, h_d^*\}$, then stop.

5 Simulation Results and Comparison with Real-Coded GA

5.1 Simulation Result Based on the Modified AS Algorithm

In order to examine the performance of the proposed modified AS algorithm, for the example given in Fig.1, a simulation experiment was executed on a personal computer with 2.66-GHz CPU and 256-MB RAM. The parameters were set as m=10, α=1, β=5, ρ=0.7, τ_0=0.01, Q=1000, and NC=200. The result is shown in Fig.5, in

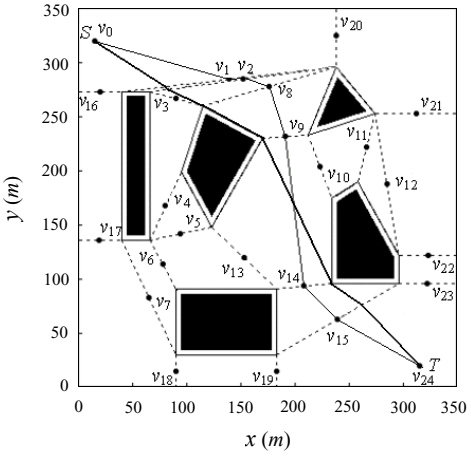


Fig. 5. Computer simulation result using the modified AS algorithm

which, the thin solid line denotes the sub-optimal path with the length of 485.851 meters, which was obtained from the Dijkstra algorithm; the thick solid line denotes the global optimal path with the length of 439.538 meters, which was obtained from the modified AS algorithm. The optimal h parameter set for this example is $\{h_1^*, h_2^*, h_3^*, h_4^*, h_5^*, h_6^*\}=\{0.2, 0.1, 0.0, 0.0, 1.0, 0.7\}$.

5.2 Comparison of Modified AS Algorithm and Real-Coded GA

We also performed a simulation experiment for the given example using the real-coded genetic algorithm (GA) with the Elitism scheme [11]. In the real-coded GA, an individual was coded as $h_1 \ldots h_i \ldots h_d$, where $h_i \in [0, 1]$, $i=1, 2, \ldots, d$. The parameters of the real-coded GA were specified as: population size m=50, crossover rate P_c=0.3, mute rate P_m=0.05, and crossover coefficient a=0.5. Fig. 6 and Fig. 7 show the convergence processes of the optimal solutions generated respectively by the two algorithms. Obviously, the modified AS algorithm has much faster convergence speed.

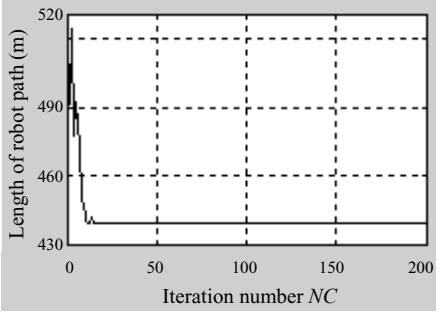


Fig. 6. Convergence tendency of the modified AS algorithm

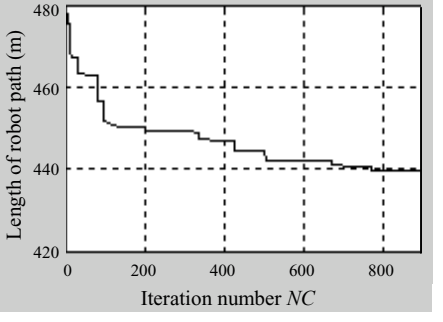


Fig. 7. Convergence tendency of the real-coded GA

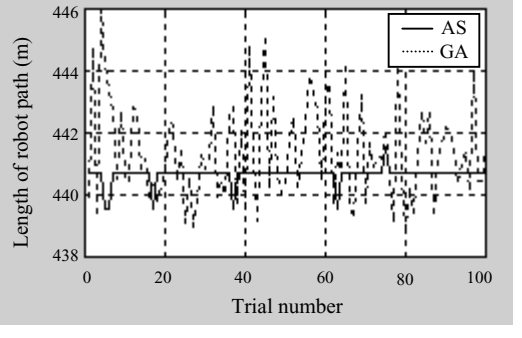


Fig. 8. Comparison of optimal solution variation of both the methods

In order to observe the variation in optimal solutions, 100 simulation trials were performed for the two methods with different random number. The result is shown in Fig. 8. From this figure, it can be seen that the optimal solution variation of the modified AS algorithm is much smaller than that of the real-coded GA.

We also adopted two statistic indexes, i.e., the mean value μ and the standard deviation σ, to evaluate the dynamic convergence behavior of the modified AS algorithm and the real-coded GA. The two statistic indexes are defined as follows:

$$\mu = \frac{\sum_{k=1}^{m} L_k}{m} \quad , \tag{10}$$

$$\sigma = \sqrt{\frac{1}{m}\sum_{k=1}^{m}(L_k - \mu)^2} \, , \tag{11}$$

where L_k represents the solution (the path length) generated by the ith individual after the modified AS algorithm or the real-coded GA completes an iteration. The less the μ value, the better the solution generated by each individual. So, μ reflects the accuracy of an algorithm. The less the σ value, the better the centrality of the solutions generated by all individuals. So, σ reflects the convergence speed of an algorithm.

Fig. 9 and Fig. 10 show the convergence tendency of μ and σ of the two methods. As can be seen, (1) after about the tenth iteration, the mean value μ of the modified AS algorithm is always smaller than that of the real-coded GA, which indicates that the solution generated by the modified AS algorithm is always better than that of the real-coded GA after the tenth iteration; (2) the mean value μ of the modified AS algorithm becomes very stable after about the twentieth iteration, whereas the mean value μ of the real-coded GA still oscillates violently after achieving the 200th iteration, which indicates that the convergence speed of the modified AS algorithm is faster than that of the real-coded GA; (3) The standard deviation σ of the modified AS algorithm is always smaller than that of the real-coded GA, which indicates that the solutions generated by the modified AS algorithm in each iteration has always better centrality compared with the real-coded GA.

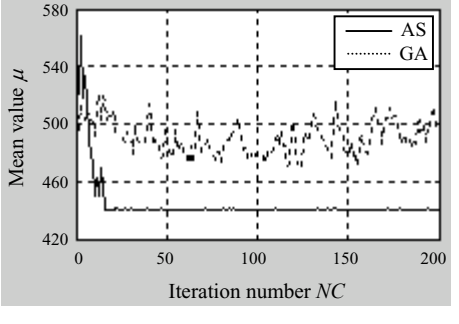


Fig. 9. Convergence tendency of mean value μ using both the methods

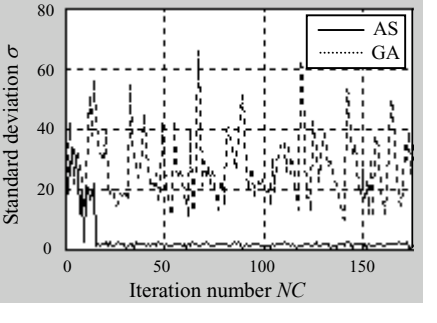


Fig. 10. Convergence tendency of standard deviation σ using both the methods

Table 1 shows the comparison data in computation efficiency of the two methods. As can be seen, the computation efficiency of the modified AS algorithm is much higher than that of the real-coded GA.

Table 1. Comparison of computation efficiency of both the methods

Methods	Average CPU time per iteration (sec.)	Average number of iterations for convergence	Average CPU time for obtaining optimal solution (sec.)
Modified AS algorithm	0.00125	56	0.07
Real-coded GA	0.00067	864	0.5789

6 Conclusions

This paper presents a novel global optimal path planning method for mobile robots based on the modified AS algorithm. The simulation results show that the proposed method is effective and the search time needed for obtaining the global optimal path is just several tens of milliseconds, which indicates that this method can be used in the real-time path planning of mobile robots. As compared with the path planning method based on the real-coded genetic algorithm, it has been confirmed that the proposed modified AS algorithm has better performance in convergence speed, solution variation, dynamic convergence behavior, and computation efficiency.

Acknowledgement

The authors gratefully acknowledge the financial support from the National Natural Science Foundation of China under Grant No.50275150 and the National Research Foundation for the Doctoral Program of Higher Education of China under Grant No. 20040533035.

References

1. Ge, S. S., Cui, Y. J.: New Potential Functions for Mobile Robot Path Planning. IEEE Transactions on Robotics and Automation, Vol. 16, No. 5, (2000) 615-620
2. Boschian, V., Pruski, A.: Grid Modeling of Robot Cells: a Memory-Efficient Approach. Journal of Intelligent and Robotic Systems, Vol. 8, No.2, (1993) 201-223
3. Yao, X. (ed.): Evolutionary Computation: Theory and Applications. World Scientific Publishing Co., Singapore (1999)
4. Tan, K.C., Lim, M.H., Yao, X., Wang L.P. (eds.): Recent Advances in Simulated Evolution And Learning. World Scientific Publishing Co., Singapore (2004)
5. Lebedev, D.: Neural Network Model for Robot Path Planning in Dynamically Changing Environment. Modeling and Analysis of Information Systems, Vol. 18, No. 1, (2001) 12-18

6. Liu, C.H., Hu, J.Q., Qi, X.N.: Path Design of Robot with Continuous Space Based on Hybrid Genetic Algorithm. Journal of Wuhan University of Technology, Vol. 27, No. 6, (2003) 819-821 (in Chinese)
7. Yung, N.H.C., Cang, Y.: An Intelligent Mobile Vehicle Navigator Based on Fuzzy Logic and Reinforcement Learning. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol.29, No.2, (1999) 314-321
8. Dorigo, M., Bonabeau, E., Theraulaz, G.: Ant Algorithms and Stigmergy. Future Generation Computer Systems, Vol. 16, No. 8, (2000) 851-871
9. Habib, M. K., Asama, H.: Efficient Method to Generate Collision Free Paths for Autonomous Mobile Robot Based on New Free Space Structuring Approach. Proceedings of the IEEE/RSJ Int. Workshop on Intelligent Robots and Systems, Osaka, Japan, (1991) 563-567
10. Dorigo, M., Gambardella, L. M.: Ant Colony System: a Cooperative Learning Approach to the Traveling Salesman Problem. IEEE Transactions on Evolutionary Computation, Vol. 1, No. 1, (1997) 53-66
11. Krohling, R.A., Jaschek, H., and Rey, J.P.: Designing PI/PID Controller for a Motion Control System Based on Genetic Algorithm. Proceedings of the 12th IEEE International Symposium on Intelligent Control, Istanbul, Turkey, (1997) 125–130

A Route System Based on Ant Colony for Coarse-Grain Reconfigurable Architecture

Song Li-Guo[1,2] and Jiang Yu-Xian[1]

[1] Department of Automatic Control,
Beijing University of aeronautics and astronautics, Beijing 100083
[2] Beijing Microelectronics Technology Institute, Beijing 100076
songlg123456@163.com

Abstract. It is very important to design a good routing-system for the whole compile-synthesis system of reconfigurable architecture (RA). Because the routing resources of coarse-grain RA (CGRA) are less than those of fine-grain RA, and several functions are often defined in same one element of RA, it is difficult to find a good route. Therefore, it is more important for routing-algorithm of CGRA to have stronger ability of finding feasible and optimum path. In the paper, the improved max—min Ant System (MMAS) that added the ability of smell for ant is applied for the routing problem of CGRA. By several benchmarks on CTaiJi that is a new developed CGRA, The improved MMAS shows better ability to find the best solution than PathFinder that is often used now.

1 Introduction

Reconfigurable Architecture (RA) is emerging as an important new computing platform for implementing intensive computations. It is well known that RA is a bridge between the flexibility of software and performance achievable in hardware [1]. Based on the width of data-path, RA can be divided into two parts: one is fine-grain RA where the typical production is FPGA (field programmable gate array); another is coarse-grain RA (CGRA) whose width of data-path is larger than one. Some mapping algorithms, which are well known for FPGA, can often be used directly for CGRA. These projects of CGRA include XPP64 [2], Morsys [3], Rapid [4] and CTaiJi [5], etc. The whole mapping technology includes two steps: place and route. Contrary to FPGA, the routing resource of CGRA is less and it is more important to design a good route system to find a feasible path than that of FPGA for the whole compile-synthesis system.

Routing problem (RP) for RA is the process of deciding exactly which routing resources will be used to carry signals from where they are generated to where they are used. Unlike many other technologies, RA has prefabricated routing resources. Thus, instead of trying to limit the size of routing channels (which is the goal of routing in standard cell), a RA router must work within the framework of the architecture's resources. Thus, the router must consider the congestion of signals in a channel, and make sure that no more routes are made through the same channel.

By now, there are mainly two routing methods for RP of CGRA. The first method is shortest-path [6]. It is to find the shortest path of every signal in turn, once the

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 215 – 221, 2006.

shortest path of a signal is found, this path is deleted from the connected graph and begin to find next signal. Therefore, the routing effect of every signal has much relation to the start sequence. Many schedule methods about the sequence of signals are appeared in [7][8][9]. The second method is pathfinder [10]. This method firstly finds the shortest path of every signal and then checks if there are congestions on edges of the connected graph. If there are congestions, the weight of the edge on which congestions occur is increased. Repeat above steps, until there are no congestions on edges of the connected graph. This method is very sensitive to parameter (*hn*, *pn*), and the results maybe changed very much with different parameters. Here, *hn* is related to the history of congestion on *n* during previous iterations of the global router, and *pn* is related to the number of other signals presently using *n*.

Dorigo put Ant Colony forward by in 1991. It is a meta-heuristic algorithm. not only can put the search behavior of ant to the around of optimum solution to improve the solution's quality and accelerative the speed of convergence, but also can effectively avoid premature. The mainly different aspect of MMAS from ordinary Ant System is that in order to limit the stagnation of the search, a more direct influence on the trail limits is exerted by limiting the allowed range of the possible trail strength by some maximum and minimum trail values: $\tau_{\min}$ and $\tau_{\max}$. The result of MMAS is better than that of other ant colony algorithms by applied to TSP.

In the paper, we improve the method of MMAS and add the ability of smell for ant. Let's suppose that there are two types of pheromone trails that ant can release: optimal trail and ordinary trail. Whenever an ant finds an optimal path, it gives out optimal trail value, whereas it gives out ordinary trail value. Further more, optimal trail value is always larger than that of ordinal. At the first glance, it may seem quite unnatural. But also Ant System is only very loosely coupled to the original behavior of the ants. The good result has been gotten by using this improved MMAS to resolve the routing problem of CGRA.

The mathematics model for route-problem is introduced in the second part of paper. The route-problem can be transformed to shortest-path problem by model properly changed. The third part is mainly about the improved MMAS algorithm, the updating of pheromone trail and how to ascertain the value of maximal trail and minimal trail. The effort of MMAS for routing is compare with that of pathfinder through instance in the last part.

2 Mathematics Model of CGRA Routing-Problem

CGRA is a network structure composed of lots of processing units (PUs). After the function of PUs is defined, we need to link the used PUs by lines that have existed. The routing-problem is the process of determining how to link these PUs. There are some requirements about the process: the delay of line should not exceed the given scale; the length sum of line should as short as possible and the congestion should not occur (i.e., the same line should not be used more than once).

If vertex v_i represents PU and direction edge e_i represents the line between PUs, the connected graph is expressed as follows:

$$G=(V, E)$$

Here V=$\{v_1, v_2, v_3 \ldots v_n\}$ is set of vertexes; E=$\{e_1, e_2, e_3 \cdots e_m\}$ is set of direction edge. $v_{s1}, v_{s2}, v_{s3} \cdots v_{sq}, v_{d1}, v_{d2}, v_{d3} \cdots v_{dq}$ are elements of V.

Let L_{sdi} (i=1,2,$\cdots q$) represents the path from v_{si} to v_{di}, it is a directed chain composed of lots of directed edges connected end to end. Then the routing problem can be defined as:

$$L_{\min} = MIN(\sum_{i=1}^{q} L_{sdi}) \tag{1}$$

The restriction conditions are:

①$\forall i$, $D(L_{sdi}) < D_{\max}$ （$D_{\max}$:maximal delay permitted）;

②$\forall i, j(i \neq j)$ $L_{sdi} \cap L_{sdj} = \phi$

These two conditions are necessary for routing-problem: the first condition guarantees that the delay of critical path is smaller than maximal delay permitted; the second condition guarantees that no congestion occurs.

3 MMAS Algorithm for Routing-Problem

It can be seen from formula 1 that the essence of routing-problem model is to look for the shortest-path. As an NP-complete problem [12], MMAS is a good choice to get the solution. A simple routing-problem is shown in figure 1, S1,S2,S3,A,B,C,D1,D2,D3 are vertexes. There are directed edges between vertexes and the weight of directed edges is shown in figure. The problem is that how ant chooses path in order to get the minimum length sum of line from S1 to D1, line from S2 to D2 and line from S3 to D3. Furthermore, no congestion occurs at the same time.

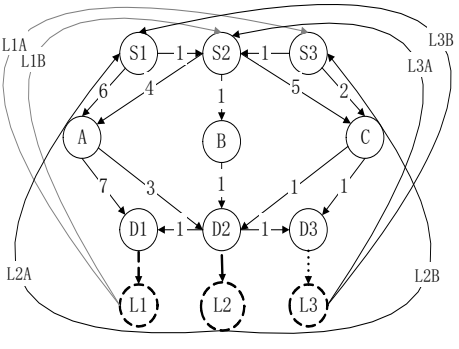


Fig. 1. Simple routing-problem

3.1 Improved MMAS Algorithm

(1) Add vertex L_i for every link pair ($S_i \rightarrow D_i$). For vertex L_i, there are one input vertex D_i and ($N_{link} - 1$) output vertexes, where N_{link} is the number of link pairs. These output vertexes point to the source vertex $S_j (j \neq i)$ of other link pair, like L1A,L1B,L2B etc in figure 1.

(2) Put Ant_k on the source vertex S_i for every link pair $S_i \rightarrow D_i$.

Ant_k moves toward destine vertex D_j by the action of pheromone trails and attractiveness.

(3) If the length that ant has moved for finding path are more than the maximum permitted, this indicate the path that ant has chosen could not be passed through and go to step 5.

(4) If Ant_k reach to the destine vertex D_j, then it directly go into vertex L_j. Only Ant_i that start from S_i can pass through vertex L_i. This ant chooses path from additional direction edges such as: L1A, L1B, L2B, by equal probability and begins to go through other link pair that has not been. If all of the link pairs have been passed through, go to next step. Otherwise, go back to step 3.

(5) The pheromone trails update until all ants finish their route. Pheromone trail only adds to optimum route and other path's pheromone trail will evaporate to decrease. Jump to step 2, and restart until the repeated times come to required number.

In order to accelerate the speed of convergence, the MMAS is improved, i.e. ant has the ability of smell. Thus, there are two kinds of pheromone trail for every edge in graph *G*:

τ_{op-ij}: pheromone trail of optimum path; τ_{ge-ij}: pheromone trail of ordinary path.

The ant needs to check the pheromone trail of edges that link to the vertex where the ant lies in before it determines the direction of next step. If the checked edge belongs to optimum path, it uses τ_{op-ij}. While if the checked edge doesn't belong to optimum path, τ_{ge-ij} is used. For edges belong to optimum path, τ_{op-ij} is always larger than τ_{ge-ij}. Therefore, the chosen path always tends to the existed optimum path. This ensure the quickly convergence of algorithm.

3.2 Updating of Pheromone Trails

The rules for pheromone trail's updating are critical for ant colony. For the improved MMAS, ant adds the smell ability. There are two kinds of pheromone trail τ_{op-ij} and τ_{ge-ij} need to update, as in formula (2) and (3).

$$\tau_{op-ij}(t+1)=\begin{cases}\rho\tau_{op-ij}(t)+\Delta\tau^{best_{ij}} & ij\in optimum-path,\ ij_{last}\in optimum-path\\ Init.Information+\Delta\tau^{best_{ij}} & ij\in optimum-path,\ ij_{last}\notin optimum-path\\ \rho\tau_{op-ij}(t) & ij\notin optimum-path\end{cases} \tag{2}$$

$$\tau_{ge-ij}(t+1)=\begin{cases}\rho\tau_{ge-ij}(t) & ij\in optimum-path,\ ij_{last}\in optimum-path\\ Init.Information & ij\in optimum-path,\ ij_{last}\notin optimum-path\\ \tau_{op-ij}(t) & ij\notin optimum-path\end{cases} \tag{3}$$

Here $ij\in optimum-path$ indicates that edge ij belongs to optimum path. $j_{last}\notin optimum-path$ indicates that edge ij doesn't belong to optimum path last time.

3.3 Boundary of Pheromone Trail

It is very important to ascertain pheromone trail scope for MMAS. In MMAS algorithm, the maximal and minimal values of pheromone trail are limited and can be express as: $\tau_{\min}\le\tau_{ij}(t)\le\tau_{\max}$. This can effectively avoid searching path falling into standstill. But, how to ascertain the value of $\tau_{\min}$ and $\tau_{\max}$?

Definition 1. When the path composed of edges chosen by maximal pheromone trail is same as the optimum path, it is defined as **MMAS convergence.**

Lemma 1. $\tau_{\max}=\dfrac{1}{1-\rho}\dfrac{1}{f(s^{opt})}$

Here ρ is evaporation parameter of pheromone trail; $f(s^{opt})$ is the value of globe optimum solution.

Lemma 2. $\tau_{\min}=\dfrac{P_{TB}\times\tau_{\max}}{(n-P_{TB})\times Avg_{\min}}$

Here P_{TB} is the mutation probability in MMAS; $Avg_{\min}$ is average of edge number on all vertexes and the number of edge does not include edge whose pheromone trail is equal to $\tau_{\max}$.

We omit the proofs of these lemmas because they are obviously true from the definition of MMAS.

4 Algorithms Compare

Because there are no standard routing test programs for CGRA, we map net lists of several digital signal-processing algorithms to CGRA. MMAS and pathfinder

algorithm is individually applied to get routing results. The CGRA we chosen is CTaiJi, it is a mesh architecture for data stream processing. It is composed of 128 16bits PUs. The test results are listed in table 1.

The values of MMAS in table 1 are statistic from the results of 40 repeated times These four examples stand for routing problem from simple to complex. For 16 points DCT, the number of operas that mapped onto architecture is minimum and the relationship of link is simple, MMAS and pathfinder both can get result very quickly. While for 16point FFT, the resources that needed are more than others and the relationship of link is complex, pathfinder do not find a feasible route. From the data in table 1, MMAS is effective for routing problem of CGRA. Its disadvantage is that it takes longer time. But, as routing resource of CGRA are simple and the number of PU is limited, the time taken by MMAS is permissible.

Table 1. Test result

algorithm			length	PU number	cost time *1 （S）
16points FFT	MMAS	Best result	44	16	324.6
		Worst result	48	18	308.5
		Average result	45.3	16.4	322.9
	pathfinder		/	/	not find path
two type IIR(L=3)	MMAS	Best result	34	9	31.5
		Worst result	35	9	35.3
		Average result	34.2	9.0	31.7
	pathfinder		35	10	12.7
12order 2parallelism FIR filter	MMAS	Best result	28	8	23.5
		Worst result	29	9	23.8
		Average result	28.0	8.0	23.6
	pathfinder		30	9	11.6
16points DCT	MMAS	Best result	8	4	1.8
		Worst result	8	4	1.8
		Average result	8.0	4.0	1.8
	pathfinder		8	4	0.4

*1：for MMAS， it is the time takes before algorithm get into convergence;
for pathfinder, it is the time that solution first occurs;

5 Conclusions

The improved MMAS for routing problem of CGRA is effective and has stronger ability of searching optimum path than pathfinder. This algorithm is very fit to CGRA whose resources are not too much. Using parallel computing technology and

improving the strategy of candidate set for ant colony are good ways to overcome MMAS shortcoming.

Acknowledgment

The National Natural Science Foundation of China No. 60503051 supports this research.

References

1. Reiner Hartenstein: A Decade of Reconfigurable Computing: a Visionary Retrospective; DATE 2001, Int'l Conference on Design Automation and Testing in Europe - and Exhibit, Munich, Germany, March 12-15, 2001
2. http://www.pactcorp.com, 2005-7-11
3. Amir H. Kamalizad, Chengzhi Pan, Nader Bagherzadeh, “Fast Parallel FFT on a Reconfigurable Computation Platform,” Proceedings of the 15th Symposium on Computer Architecture and High Performance Computing (SBAC-PAD’03), 2003,
4. A.K.W.Yeung, J.M.Rabaey: A Reconfigurable Data-driven Multi-Processor Architecture for Rapid Prototyping of High Throughput DSP Algorithms; Proc. ICASSP’97 Munich, Germany, April 1997
5. Song Liguo, Jang Yuxian, “CTaiJi——a new coarse grain re-configurable computing platform”, microelectronics and computer,” september,2005
6. Lee, C. Y. "An Algorithm for Path Connections and Its Applications." IRE Trans. On Electronic Computers, vol. EC-10, Sept 1961, 346-365.
7. W. Dees and R. Smith, "Performance of Interconnection Rip-Up and Reroute Strategies," in Proc. 18th Design Automation Conference, June 1981, pp. 382-390.
8. D. Hill, "A CAD System for the Design of Field Programmable Gate Arrays," in Proc. 28th Design Automation Conference, June 1991, pp. 187-192.
9. R. Linsker, "An Iterative-Improvement Penalty-Function-Driven Wire Routing System," IBM Journal of Research and Development, vol. 28,Sept. 1984, pp. 613-624.
10. L. McMurchie ,C. Ebeling “PathFinder: A Negotiation-Based Performance-Driven Router for FPGAs”
11. Li shiyong, “ant colony and its application”, published by HaErBin University of Industrial,2004
12. http://www.inst.eecs.berkeley.edu/~cs150/sp05/Lab/Lab1.PDF, 2005-7-11
13. S. Brown, J. Rose, and Z. Vranesic, "A Detailed Router for Field-Programmable Gate Arrays," IEEE Transactions on Computer-Aided Design, vol. 11, no. 5, May 1992, pp. 620-628.

Robot Planning with Artificial Potential Field Guided Ant Colony Optimization Algorithm

Dongbin Zhao and Jianqiang Yi

Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, P.R. China
{Dongbin.zhao, jianqiang.yi}@ia.ac.cn

Abstract. This paper investigates the problem of robot planning with ant colony optimization and artificial potential filed algorithms. Robot planning is to find a feasible path from a source to a goal while avoiding obstacles in configuration space. Artificial potential field (APF) is verified as an efficient method to find a path by following the maximum potential field gradient. But it suffers from the local minima. However, ant colony optimization (ACO) is characterized as powerful probabilistic search ability, which is thought to be fit for solving such local minima problems. By the combination of both merits, an APF guided ACO algorithm is proposed, which shows some good features in searching for the optimal path solution. The length optimal path solution can always be achieved with the proposed hybrid algorithm in different obstacles environment from simulation results.

1 Introduction

Robot planning is to find a feasible or optimal path from a source to a goal while avoiding obstacles in configuration space. The problem has been widely investigated, and some famous methods have been proposed, such as artificial potential field (APF) [1-6], probabilistic roadmap planner (PRM) [7-8], sensor-based method [9], *etc*. They all have their merits and shortcomings. The PRM is fit for planning the robot with multiple degrees of freedom, but hard to be extended in dynamic environment. The sensor-based method is simple but the generated path is rather zigzag. The APF is to find a path by following the maximum potential field gradient. It is easy to be applied in practical robot with the local potential field constructed by sensing the local environment with range sensors. It also facilitates the application in dynamic environment, but still suffers from the problem of local minima. Several methods, such as harmonic function [2-4], and other hybrid algorithms [5-6] were proposed to overcome such shortcoming.

Ant colony optimization (ACO) [10-13] was originated from natural ant colony behavior. It gained great successful applications for traveling salesman problem, quadratic assignment problem, scheduling problem, telecommunications networks routing problem, *etc*. As to our knowledge, there are few attentions paid to the application of ACO on path planning problem [14-15]. In our previous work [16], a modified ACO algorithm for robot path planning was proposed.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 222 – 231, 2006.

It was indicated that the modified ACO was featured as probabilistic search ability to explore the space. A feasible solution could be always achieved, but the optimal solution could not be guaranteed and the learning process was rather time consuming. However, APF will provide some deterministic information for the environment, which could be used to guide the robot from the source to the goal effectively. At the same time, the local minima shortcoming of APF could be overcome by probabilistic search ability of ACO.

2 Deterministic Planning with Artificial Potential Field

2.1 Robot Planning Problem

Robot planning is to find a feasible or optimal path from a source to a goal while avoiding obstacles in configuration space. The space can be discretized to a series of grid points according to certain sample resolution. The source and the goal are on the grid points. Obstacles are just in the path from the source to the goal, and they can be thought as sets of the grid points. In one step, the mobile robot moves from the current point to the next point, one of its eight neighbors. Then a feasible path is constructed by a series of such moving steps.

The sampling resolution is represented by d_g, the distance between two horizontal or vertical neighbor points, the robot path is updated by

$$p = p + d_g \left[sign(\cos\theta_i), sign(\sin\theta_i) \right], \theta = \left\{ \frac{\pi}{4}, \frac{\pi}{2}, \ldots, 2\pi \right\}. \tag{1}$$

where θ represents one of the eight directions. If the value of the trigonometric function is zero, then the function *sign* returns 0, otherwise, it returns 1 or –1.

2.2 Artificial Potential Field

In an artificial potential field method, the goal is represented by an attractive potential, while the source and the obstacles are represented by repulsive potentials. So the robot can reach the goal from the source while avoiding obstacles. Harmonic function, featured as free of local minima, is often adopted to construct the artificial potential field. A spherical symmetry function is used to represent the potential.

$$\phi_{s,g,o} = \lambda \log r. \tag{2}$$

where r is the distance from the current position to the goal, source, or obstacles. The subscripts s, g, and o represent source, goal, and obstacles respectively. λ represents attractive or repulsive strength.

A uniform harmonic function is denoted as

$$\phi_u = -\lambda_u \left(x\cos\theta_u + y\sin\theta_u \right). \tag{3}$$

where θ_u is angle between axis x and the direction from the source to the goal.

The whole potential is the superposition of all harmonic functions. The velocity components are calculated as the differentiation with respect to x and y.

$$v_x = -\phi_x = \sum \lambda_{s,g,o} \frac{x_{s,g,o} - x}{\left(x - x_{s,g,o}\right)^2 + \left(y - y_{s,g,o}\right)^2} + \lambda_u \cos\theta_u \tag{4}$$
$$v_y = -\phi_y = \sum \lambda_{s,g,o} \frac{y_{s,g,o} - y}{\left(x - x_{s,g,o}\right)^2 + \left(y - y_{s,g,o}\right)^2} + \lambda_u \sin\theta_u$$

In continuous configuration space, the robot path can be constructed successfully by following the velocity direction. But the continuous harmonic function should be described in a discrete form based on range sensors for the environment in practical application. Due to the error caused by discrete calculation, local minima may also exist. A discrete laplace method could be used instead, but a feasible not the optimal solution is the desired [4].

As to our knowledge, discrete harmonic potential functions are the widely adopted algorithms for obstacles avoidance. The robot will move to the next point, one of its discrete neighbors, with the maximum velocity component.

$$v = \sqrt{v_x^2 + v_y^2} \tag{5}$$
$$\theta_v = a\tan 2(v_y, v_x)$$

where θ_v is the velocity direction. By comparison, one of the eight directions closest to the velocity direction is selected as the robot moving direction.

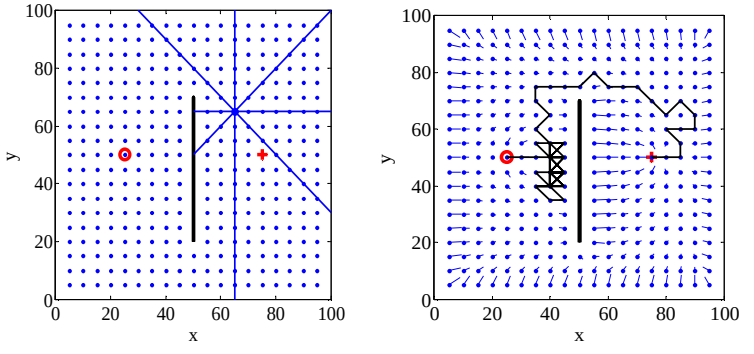


Fig. 1. Measured obstacles distances and a feasible path achieved with artificial potential field

An example of robot planning with artificial potential filed is shown in Fig.1, where a source, a goal and an obstacle are indicated as a circle, a cross and a line respectively. The blue solid circle represents one possible robot position. The measured distances in eight directions to its surrounding including the obstacle and borders are shown with blue lines in the left side of Fig.1. Then, the velocity field can be calculated with Equ.(2) to (5), and indicated as blue radials with velocity magnitude and direction in the right side of Fig.1. A feasible path, not the optimal, could be found by following the maximum velocity component from the source to the goal.

Where the robot is in local minima, a commonly used method is to select a random direction, which will cause the path rather zigzag and lengthy, shown in Fig.1.

3 Probabilistic Search with Ant Colony Optimization

As indicated in [16], we proposed a modified ACO algorithm for robot planning problem, which showed good performance on finding a feasible path with a high probability. The iteration learning process consists of the following steps.

3.1 Initialization

The current robot position p is set to the source point p_s initially. θ_i represents moving direction. $\Delta\tau$ and τ are pheromone increment and pheromone on the robot path. Supposed the pheromone on all the paths are the maximum $\tau_{\max}$ initially.

$$\Delta\tau(p,\theta_i)=0, \tau(p,\theta_i)=\tau_{\max}, i=1,2,\cdots 8. \tag{6}$$

3.2 Finding a Feasible Path

The probability p^k of the next point to be selected is

$$p^k(p,\theta_i)=\begin{cases}\dfrac{\tau(p,\theta_i)^\alpha\cdot\eta(\theta_i)^\beta}{\sum_{i=1}^{8}\tau(p,\theta_i)^\alpha\cdot\eta(\theta_i)^\beta} & if\ \ p_n\notin obs\\ 0 & else\end{cases}. \tag{7}$$

where η is a heuristic function, defined as the reciprocal of the distance from the current point p to the next point p_n. α and β are weights of pheromone and heuristic function respectively. Roulette method or other rank methods can be selected to determine the next point according to the probabilities of all candidate points.

m ants locate at the same source point. The *kth* ant path is stored in the path list $path^k$. Supposed the robot could not penetrate the obstacles, only the borders of the obstacles and the environment are considered in the set *obs*. So once a feasible path is achieved, it must be the best path from the source to the goal. Then, the searching process stops. A path pruning mechanism is followed to remove repeated points in the path to derive a clean path with the robot stepping on each point once [16]. The global best path is also obtained by comparing the best path with the previous best values.

3.3 Update of Pheromone Increment

The pheromone increment is calculated by

$$\Delta\tau(p,\theta_i)=\begin{cases}\dfrac{Q}{L^{best}} & if\ \ p\in path^k\\ 0 & else\end{cases}, i=1,2,\cdots 8. \tag{8}$$

where L^{best} is the length of the best path in each iteration step. Q represents the amount of the pheromone quantity deposited on the best path. The shorter the path is, the higher the pheromone increment on the path is. The global best solution is used to replace the best path of the current iteration step once for many iteration steps. This assures the iteration converge to the optimal solution with a high probability.

3.4 Update of Pheromone

Considering evaporation effect on the path, the pheromone is updated as

$$\tau(p,\theta_i) = (1-\rho)\tau(p,\theta_i) + \Delta\tau(p,\theta_i), i = 1,2,\cdots 8 . \tag{9}$$

where evaporation coefficient ρ is an constant from 0 to 1, to represent the pheromone decreasing rate. The updated pheromone is further limited in a scope [$\tau_{\min}$, $\tau_{\max}$]. To maintain a large pheromone on the best path, an evaporation related pheromone quantity mechanism is adopted, denoted by

$$Q \approx \rho L^{estimate} \tau_{\max} . \tag{10}$$

where $L^{estimate}$ is an estimate value of the best path length. The ant number can also be assumed as $L^{estimate}$ divided by d_g.

3.5 Stop Criterion

If the iteration process reaches the maximum step, it stops and the final optimal path outputs. Otherwise, the algorithm returns to the part of finding a feasible path and continue the iteration process.

4 APF Guided ACO Algorithm

As mentioned in section 2, artificial potential field provides deterministic information about the configuration space, which facilitates the search process of path planning. But the robot may be trapped in local minima in discrete configuration space even with the harmonic potential function. The path is not with the minimum length, but to follow the maximum velocity gradient. However, ACO is indicated as powerful probabilistic search ability to escape the local minima. The positive feedback mechanism will guarantee the optimal solution with a high probability.

To combine the merits of both algorithms, a hybrid algorithm is presented. The path planning is conducted under the infrastructure of ACO. But the APF is introduced in the step of find a feasible path, to guide the searching process efficiently. Then, Equ.(7) in ACO is modified as

$$p^k(p,\theta_i) = \begin{cases} \dfrac{\tau(p,\theta_i)^{\alpha} \cdot \eta(\theta_i)^{\beta} \cdot \sigma(p,\theta_i)^{\gamma}}{\sum_{i=1}^{8} \tau(p,\theta_i)^{\alpha} \cdot \eta(\theta_i)^{\beta} \cdot \sigma(p,\theta_i)^{\gamma}} & if\ p_n \notin obs \\ 0 & else \end{cases} . \tag{11}$$

where σ is the velocity function calculated by APF, and γ is the weight. The velocity component on each one of the eight directions and the weight are calculated by

$$\begin{aligned} \sigma(p,\theta_i) &= \exp(\cos(\theta_v - \theta_i)) \\ \gamma &= \lambda_v v \end{aligned} \quad . \tag{12}$$

which means the velocity component with the direction closest to the velocity direction will get the maximum value. The velocity value is proportional to the weight.

From Equ.(11) and (12), we can see that if the point is in local minima with a zero velocity value, the velocity function item in Equ.(11) will be eliminated completely. That means the algorithm turns to the single ACO. The robot path will be found by the powerful probabilistic search ability free of the problem of local minima.

If the velocity value is large, the velocity function will have a more important influence on determining the robot path, so that the influence of probabilistic search property by ACO is decreased. The deterministic velocity value will facilitate the path planning process.

Equ.(4) indicates that several attractive and repulsive strength coefficients affect the velocity. By strengthening obstacles repulsive strength, a safety path solution is inclined to achieve. On the other hand, a length optimal path with less safety may be obtained instead. This is verified in the following simulation results.

5 Simulation and Comparison

The configuration space is with the dimensions of [100 100] points, which is sampled with grid distance of 5 points. The obstacles are the same with [16] to test and compare the algorithm performances. The maximum iteration step is 500. The global best solution is used to replace the best solution every 50 steps. The parameters of the APF guided ACO algorithm are tabulated in Table 1.

Table 1. Parameters of the APF guided ACO algorithm

λ_s	λ_g	λ_o	λ_u	λ_v	α	β	$\tau_{\max}$	$\tau_{\min}$	ρ
-10	10	-10	1	0.1	1	1	1	0.01	0.05

5.1 Single Simple Obstacle

The problem is shown in Fig.2(a). The source and the goal of the robot are [25,50] and [75,50], indicated as a circle and a cross respectively. An asymmetrical line obstacle is from [50,20] to [50,70], shown as a black line. There is only one optimal path solution from the source to the goal while avoiding the obstacle.

The path length is estimated to be 100, so the ant number *m* is calculated to be 20. Ten experiments are carried out. A promising result is achieved that the optimal solution can be always achieved, compared to the probability of 70-80% in [16]. One

example is shown in Fig.2, where the left shows the final pheromone intensity distribution, proportional to the line width, and the right is the best path length. About step 100, the path length converges to the optimal solution. The convergence steps are shorter than about 130 steps in [16].

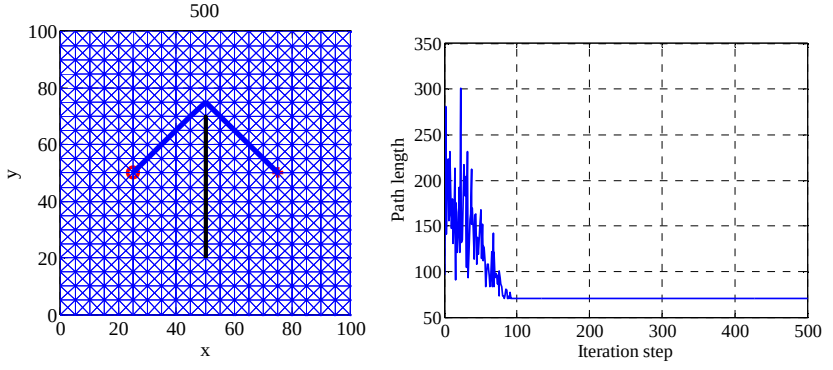


Fig. 2. Robot planning in a single simple obstacle environment. (a) Pheromone intensity distribution at iteration step 500; (b) The best path length.

5.2 Single Complex Obstacle

The problem is shown in Fig.3. The source and the goal of the robot are [40,50] and [80,50] respectively. A rather complex symmetrical obstacle is shown in black part. There are many optimal path solutions from the source to the goal while avoiding the obstacle, for example, either in the upper or lower blank area. The experiment is considered to test the path planning performance of the algorithm to find one of the optimal paths in a difficult environment.

The path length is estimated to be 150, so the ant number m is calculated to be 30. Simulation results do not indicate better performance than the single ACO algorithm [16]. By analyzing the velocity field and one sub-optimal solution, shown in Fig.3(a) and (b), the robot path or direction is affected by a large velocity magnitude. The length optimal solution is to follow the path just above or below the obstacle, but where the large velocity magnitude tends to guide the robot away from the obstacle. The solution can be thought as another safety optimal solution.

To achieve the length optimal solution, it is a feasible way to the take little consideration of the path safety. The obstacles strength is reduced to –1.25. Ten simulation experiments are also conducted, and the result is quite different. The length optimal solution can always be achieved. The velocity field and one pheromone intensity distribution at step 500 are shown in Fig.3(c) and (d). Compared to [16], the probability of the optimal solution is effectively increased. The result indicates that the obstacle strength plays an important role in determining a length optimal or safety optimal solution. Furthermore, there is much less fluctuations in the best path length, shown in Fig.3(e).

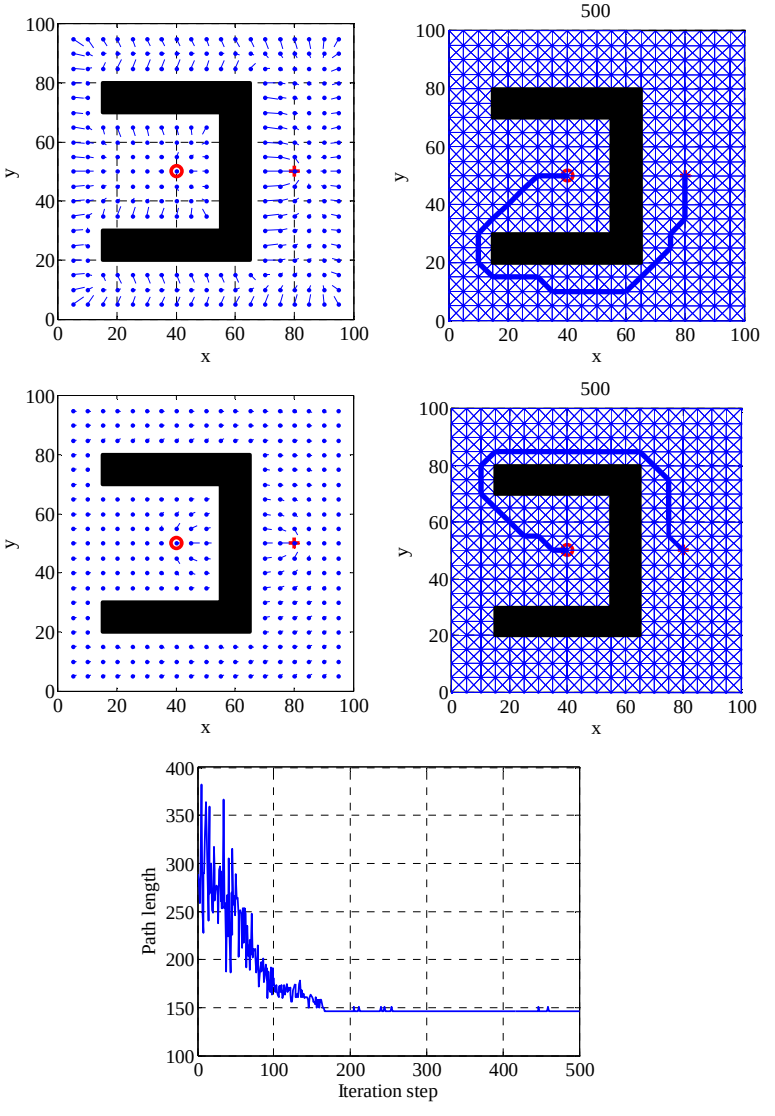


Fig. 3. Robot planning in a single complex obstacle environment. (a) Velocity field; (b) Pheromone intensity distribution at iteration step 500; (c) Velocity field with a reduced obstacle repulsive strength; (d) Pheromone intensity distribution at iteration step 500 with a reduced obstacle repulsive strength; (e) The best path length with a reduced obstacle repulsive strength.

5.3 Multiple Obstacles

The problem is shown in Fig.4. The source and the goal of the robot are [10,10] and [90,90] respectively. Multiple rectangle obstacles with different width and height are distributed in the environment randomly, shown in black parts. There are several

optimal path solutions from the source to the goal while avoiding the obstacles. The experiment is considered to test the path planning performance of the algorithm with more obstacles and narrow passage, a typical problem of robot planning.

As indicated in single ACO algorithm [16], the optimal solution can always be achieved. This is because the narrow passage reduces the searching space greatly so that the problem difficulty is reduced. The proposed hybrid algorithm also shows the same performance of the optimal solution achievement, but the convergence time is quite shorter than that of [16], shown in Fig.4.

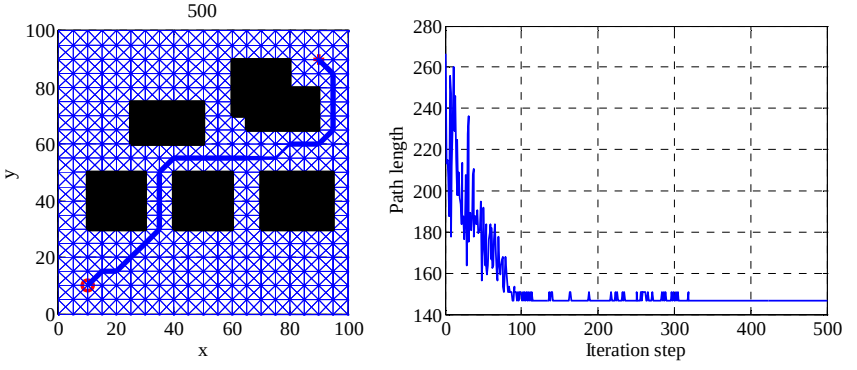


Fig. 4. Robot planning in multiple obstacles environment. (a) Pheromone intensity distribution at iteration step 500; (b) The best path length.

6 Conclusions

By combining the merits of deterministic planning of artificial potential field (APF), and probabilistic search ability of ant colony optimization (ACO), a hybrid algorithm is proposed for the problem of robot planning. Simulation results indicate better performances including iteration convergence time and the probability of the length optimal solution than that of single APF or ACO algorithm. The idea of incorporating deterministic information in ACO could also be applied in other application fields.

Acknowledgement

This work is funded partly by Natural Science Foundation of China No.60475030, Most Project of China No.2004DFB02100, international cooperative project on intelligence and security informatics by the Chinese Academy of Sciences, and Joint Laboratory of Intelligent Sciences & Technology of Chinese Academy of Sciences.

References

1. Khatib, O.: Real-time Obstacle Avoidance for Manipulators and Mobile Robots, The Int. J. of Robotics Research 5(1) (1986) 90-98
2. Rimon, E., Koditschek, D.E.: Exact Robot Navigation Using Artificial Potential Functions, IEEE Trans. on Robotics and Automation 8(5) (1992) 501-517

3. Kim, J.O., Khosla, P.K.: Real-time Obstacles Avoidance Using Harmonic Potential Function. IEEE Trans. On Robotics and Automation 8(3) (1992) 338-349
4. Connolly, C. I., Burns, J. B., Weiss, R.: Path Planning Using Laplace's Equation. Proc. of the IEEE Int. Conf. on Robotics and Automation Vol.3 (1990) 2102-2106
5. Rosell, J., Iniguez, P.: Path planning using Harmonic Functions and Probabilistic Cell Decomposition. Proc. of the IEEE Int. Conf. on Robotics and Automation (2005) 1803-1808
6. Kazemi, M., Mehrandezh, M.: Robotic Navigation Using Harmonic Function-based Probabilistic Roadmaps. Proc. of the IEEE Int. Conf. on Robotics and Automation (2004) 4765-4770
7. Kavraki, L.E., Svestka, P., Latombe, J.C., Overmars, M.H.: Probabilistic Roadmaps for Path Planning in High-dimensional Configuration Spaces, IEEE Trans. on Robotics and Automation 12(4) (1996) 566-580
8. Nielsen, C.L., Kavraki, L.E.: A Two Level Fuzzy PRM for Manipulation Planning, Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (2000) 1716-1721
9. Lumelsky, V.J., Stepanov, A.A.: Dynamic Path Planning for A Mobile Automaton with Limited Information on The Environment, IEEE Trans. on Automatic Control AC-31(11) (1986) 1058-1063
10. Dorigo, M., Maniezzo, V., Colorni, A.: Positive Feedback as A Search Strategy, Technical Report (1991)
11. Dorigo M., Gambardella, L.M.: Ant Colony System: A Cooperative Learning Approach to The Traveling Salesman Problem, IEEE Trans. on Evolutionary Computation 1(1) (1997) 53-66
12. Stützle, T., Hoos, H.: MAX-MIN Ant System and Local Search of The Traveling Salesman Problem, Proc. of the IEEE Int. Conf. on Evolutionary Computation (1997) 309-314
13. Stützle, T., Dorigo, M.: ACO Algorithms for The Traveling Salesman Problem. In: Miettinen, K., Makela, M., Neittaanmaki, P., Periaux J. (eds): Evolutionary Algorithms in Engineering and Computer Science. Wiley (1999)
14. Fan, X.P., Luo, X., Yi, S., Yang, S.Y., Zhang, H.: Optimal Path Planning for Mobile Robots Based on Intensified Ant Colony Optimization Algorithm, Proc. of the Int. Conf. on Robotics, Intelligent Systems and Signal Processing (2003) 131-136
15. Hu, X.B., Huang, X.Y.: Path Planning in 3-D Space for Robot Based on Ant Colony Algorithm, J. of Chongqing University 27(8) (2004) 132-135
16. Zhao, D.B., Yi, J.Q.: A Framework of Mobile Robot Path Planning with Ant Colony Optimization. Proc. of the Int. Conf. on Sensing, Computing and Automation (2006) 2009-2013

Heuristic Searching Algorithm for Design Structurally Perfect Reconstruction Low Complex Filter Banks*

Zhe Liu[1] and Guangming Shi[2]

[1] School of Science, Northwestern Polytechnical University,
Xi'an, 710072, China
liuzhe@nwpu.edu.cn
[2] School of Electronic Engineering, Xidian University, Xi'an,710071, China
gmshi@xidian.edn.cn

Abstract. Filter banks are found many applications for images and signal processing. A lower complex filter banks are desired for their effective implementation. In this paper, we address a problem how to design low complex filter banks. A perfect restructure (PR) filter banks can be factorized by lifting steps and the procedure of factorization are described as multi-fork tree-structure. A heuristic-searching algorithm was proposed for finding the optimization PR filter banks with low dynamical coefficients. The given examples show the proposed method is effective.

1 Introduction

Filter banks are popular in the applications various speech and image processing applications. Recently, there is increasing interest in designing filter banks with low implementation complexity because of the demand for signal processing in real time. At same time, a structurally PR FBs because the system can be implemented by using sum of powers-of-two (SOPOT) [1] coefficients are popular. With the use of SOPOT, coefficient multiplications can be implemented with simple shifts and additions. The complex of implementation and calculations of filter banks is reduced.

Unfortunately, due to the large dynamic range of coefficients, frequency responses of the FBs with SOPOT coefficients are not as good as FBs with infinite precision arithmetic if the shorter terms of SOPOT are chosen. Therefore, the FBs with coefficients in lower dynamical range are preferred. The iterative searching method [2] for solving the problem is very complex because there exist many optimization calculations.

In this paper, a novel a heuristic-searching algorithm is proposed for solving the problems. Firstly, by using a lifting factorization algorithms [3], a non-structurally PR or nearly PR FBs can be converted into a structurally PR FBs. Two kinds of long division are used in the factorization [4]. Because of non-unique factorization results

* This work has been supported by "Excellent young teacher fund" in NWPU, the National NSF of China (Grant No. 60372047) and. NSF of Shaanxi Province 2005F44, China.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 232 – 235, 2006.

by long division, whole procedures of possible lifting factorizations are described as multi-fork tree-structure. There exist many factorization paths for getting a group of the coefficients of filter bank. During the group of coefficients, there must be one group with minimal dynamic range coefficients. Secondly, a heuristic-searching algorithm is employed to find the best factorization path in order to get coefficients in low dynamical range.

2 Factorization of Two or M-Channel FIR FBs

The PR condition [5] of two channel filter bank with analysis filters $\{ H_0(z),\ H_1(z) \}$ and synthesis filter $\{ F_0(z),\ F_1(z) \}$ can be expressed as following when setting $F_0(z) = -H_1(-z)$ and $F_1(z) = H_0(-z)$:

$$H_{00}(z)H_{11}(z) - H_{01}(z)H_{10}(z) = \beta \cdot z^{-d}, \tag{1}$$

where $H_0(z) = H_{00}(z^2) + z^{-1}H_{01}(z^2)$, $H_1(z) = H_{10}(z^2) + z^{-1}H_{11}(z^2)$, β is some constant and d is some integer.

And the PR condition of *M*-channel maximally decimated filter banks constructed by a prototype filter $h(n)$ with Cosine-Modulated [5] is given as following:

$$G_k(z)G_{2M-1-k}(z) + G_{M+k}(z)G_{M-1-k}(z) = cz^{-d}, \tag{2}$$

where $H(z) = \sum_{k=0}^{2M-1} z^{-k}G_k(z^{2M})$, $0 \le k \le M-1$, and d ($0 < d \le 2m-1$) is a delay parameter of the filter banks. We can find that Eqn.(1) and Eqn.(2) have same format.

Using lifting factorization [3][4] on Eqn.(1) and (2), we have

$$\begin{bmatrix} H_{10}(z) & H_{00}(z) \\ H_{11}(z) & H_{01}(z) \end{bmatrix} = \prod_{i=1}^{n} \begin{bmatrix} q_i(z) & z^{-k_i} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} K & -Q(z) \\ 0 & (-1)^{n-1}\beta / K \end{bmatrix}, \tag{3}$$

$$\begin{bmatrix} G_k(z) & G_{M-1-k}(z) \\ G_{M+k}(z) & G_{2M-1-k}(z) \end{bmatrix} = \prod_{i=1}^{j} \begin{bmatrix} q^{(k)}{}_i(z) & z^{-k_i} \\ 1 & 0 \end{bmatrix} \prod_{i=j+1}^{n} \begin{bmatrix} q^{(k)}{}_i(z) & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} K^{(k)} & -Q^{(k)}(z) \\ 0 & (-1)^{n-1}\beta / K^{(k)} \end{bmatrix}. \tag{4}$$

Where $q_i(z)$ and $Q(z)$ are polynomial. There are two kinds of long division are used in the lifting factorization and factorization results are non-unique [4]. Because of constrained by PR conditions, number (j) of $q_i(z)$ terms of fore-long division should satisfied [4]

$$\sum_{i=1}^{\mathbf{j}} (|\, q_i(\mathbf{z}) \,| + 1) = d, \tag{5}$$

where $|\ |$ is length operation. We find that if the order of fore- or back-division to be taken is arbitrarily arranged during the procedures of factorization, $q_i(z)$ will be in large dynamical range.

3 Heuristic Searching Algorithm

Here we propose heuristic searching algorithm in order to get a low dynamical coefficients of filter. Assume that it needs *n* steps to finish whole factorizations. The possible paths of the factorization are like a tree-structure with multi-fork shown in Fig.1.

An evaluation function of dynamical range in each branch is defined as:

$$\Phi(i,m)=\frac{\max[q_1(z),\cdots,q_{i-1}(z),q_{i,m}(z)]}{\min[q_1(z),\cdots,q_{i-1}(z),q_{i,m}(z)]},. \tag{6}$$

Because of the constraint of Eq.(5), there are four possible strategies in some levels of the tree**.** Assume that we have processed the factorization on i -th step, $q_1(z),\cdots,q_i(z)$ are obtained, and there exist 4 branches (or 4 strategies) in the $(i+1)$ -th step. Then four evaluation functions $\Phi(i+1,m)(1\le m\le 4)$ are calculated. A branch with minimal $\Phi(i,m)$ is chosen in $(i+1)$ -th step and also let $q_{i+1}(z)=q_{1,m}(z)$, then continue next steps. The heuristic searching algorithm is formulated as follows:

(1) Give an predefined dynamical range Φ .Start the factorization from the root of the tree. There are 4 possible start points, and one of them is chosen as initiation point called father ($i=1$).
(2) Determine how many children (strategies, or *m*) exist in the next stop according to Eq.(5)
(3) Factorize $q_i^k(z)$ using the possible strategies and obtain $q_{i,p}(z)(1\le p\le m)$
(4) Calculate the evaluation functions $\Phi(i,p)$ for every $q_{i,p}(z)$
(5) Sort ascending order of branches with $\Phi(i,p)$ as first child, second child, $\cdots$.
(6) Chose the branch with minimal evaluation functions $\Phi(i,p)$

If minimal $(\Phi(i,p))\le\Phi$, then the children upgrade as fathers, continue the step (2)~(6) until the all $q_i(z)$ $(1<i<n)$ are found.

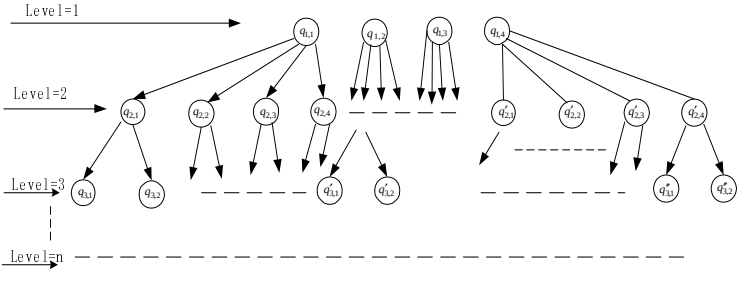


Fig. 1. Tree structure of factorization path

4 Design Examples

We now present an example which is a low-delay PR two-channel FB with length $N_0 = N_1$ =12 and d=3. The optimal factorization matrix is

$$\begin{bmatrix} H_{10}(z) & H_{00}(z) \\ H_{11}(z) & H_{01}(z) \end{bmatrix} = \prod_{i=1}^{2} \begin{bmatrix} q_i(z) & z^{-k_i} \\ 1 & 0 \end{bmatrix} \prod_{i=3}^{9} \begin{bmatrix} q_i(z) & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} K & -Q(z) \\ 0 & -\beta / K \end{bmatrix}. \tag{7}$$

Their magnitudes of minimum and maximum coefficient are 0.110361 and 3.478578, and the dynamical range of the coefficient is 31.52. Another example is 5-chaanel Cosine-Modulated filter bank with the length-30 of prototype filter and $d = 1$. If factorization path is arbitrary and the magnitude of minimum and maximum coefficient is 0.7561e-2 and 1.1995e+1, respectively, thus the dynamical range is 1.5863e+3. If the heuristic searching algorithm is used, we get an optimal factorization path. And the magnitude of minimum and maximum coefficient is 3.5330e-2 and 3.7980e-1, respectively. And the dynamical range of the coefficients is 1.0750e+1. For the space limited, the detail coefficients are not given here.

5 Conclusions

By using lifting factorization and the heuristic searching algorithm, a novel method for design M-channel structurally PR FB with lower dynamical range coefficients is presented. It is useful for reducing the complexity of implementation of FBs. The results given in the paper illustrate that the method is effective.

References

[1] Chan, S.C. et al Multiplierless perfect reconstruction modulated filter banks with sum-of-powers-of-two coefficients IEEE SP Letters, ,Vol.8 (6) (2001) 163-166

[2] Ying-Jui Chen et al. Theory and factorization for a class of structurally regular biorthogonal filter banks, IEEE Trans. on SP, Vol. 54 (2), (2006) 691–700

[3] Daubechies I.C., et al. Factoring Wavelet transform into lifting steps. J.Fourier Analy Application , (1998)247-269,

[4] Shi G.M. et al "Design of structurally PR two-channel FIR filter banks with linear phase" Chinese Journal of Electronics, vol:11(2) (2002) 1-4

[5] Vaidyanathan P. P. Multirate systems and filter banks. Englewood Cliffs, NJ: Prentice Hall, (1993).

Blind Multi-user Detection for Multi-carrier CDMA Systems with Uniform Linear Arrays*

Aifeng Ren[1,2] and Qinye Yin[1]

[1] School of Electronics and Information Engineering,
Xi'an Jiaotong University, Xi'an 710049, P.R. China
[2] School of Electronic and Engineering,
XiDian University, Xi'an 710071, P.R. China
renaifeng@xinhuanet.com, qyyin@xjtu.edu.cn

Abstract. To mitigate the inter-chip and inter-symbol interferences (ISI), multi-carrier code-division multiple access (MC-CDMA) communication systems, which integrate the advantages of multi-carrier transmission systems with those of CDMA, are considered as promising techniques for future broadband wireless multimedia communications. In this paper, we focus on the uplink multi-user detection (MUD) on the scheme that a uniform linear array (ULA) is applied to the base station of the MC-CDMA system. We first describe the equivalent spatial-temporal channel model of multi-user multiple-input multiple-output (MIMO) MC-CDMA systems. Based on this scheme, by utilizing the finite alphabet property of transmitted symbols, an multi-user detector is derived for the MC-CDMA system with the ULA over frequency-selective fading channels. Computer simulations illustrate that our algorithm is more robust against noise and can well mitigate multiple access interference (MAI) in multi-user scenarios. In particular, the proposed scheme has the potential of providing the anticipated MUD performance gains with a complexity that would be manageable for MC-CDMA systems.

1 Introduction

Future wireless communication systems need to support a very high transmission data rate in order to provide voice and multimedia services. Multi-carrier (MC) technologies are emerging as a considerable force to support the needs of next generations (NextG) of wireless communication systems. One of MC modulation schemes is orthogonal frequency division multiplexing (OFDM), in which a single high-rate data stream is transmitted over a number of lower rate subcarriers, which makes the system robust against frequency-selective fading and intersymbol interference (ISI). As a form of combining code-division multiple access (CDMA) and OFDM, the multicarrier (MC) CDMA scheme was originally proposed by N. Yee et al in PIRMC'93 for high-data-rate wireless multimedia applications[1]. Owing to merits of the efficiency

* This work was supported in part by the National Natural Sciences Foundation (No. 60572046 and No. 60502022) and the Research Fund for Doctoral Program of Higher Education (No. 20030698027) of China.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 236 – 244, 2006.

to utilize the available spectrum and the capability to meet the demand of diverse multimedia services, MC-CDMA techniques are strong contestants for NextG wireless multimedia communications. Different variations of combination of CDMA and OFDM are known under acronyms such as MC-CDMA, MC-DS-CDMA (multicarrier direct sequence CDMA), and MT-CDMA (multitone CDMA)[2]. The MC-CDMA system we consider in our study is the one where signal is spread and then converted into parallel data stream to be transmitted through different subcarriers.

The MC-CDMA scheme is a promising technique for relieving capacity limit problems of conventional CDMA systems. MC-CDMA signals do not experience significant linear distortion in frequency-selective fading channels because the symbol duration is much longer than the delay spread. Another different approach to further increase the system capacity is the use of spatial processing technique with antenna arrays at the base station of wireless communication systems.

2 Model of Multi-User MIMO MC-CDMA

Here, we assume that the base station ULA with M_R elements is employed and assumed that K users share the same set of subcarriers. The number of subcarriers equals to the length of spreading code G. Fig.1 shows the baseband model of the multi-input multi-output (MIMO) MC-CDMA systems. "IDFT" shown in Fig.1 is the corresponding all-digital implementation of the MC modulation bank, while "DFT" is the corresponding all-digital implementation of the MC demodulation bank [3].

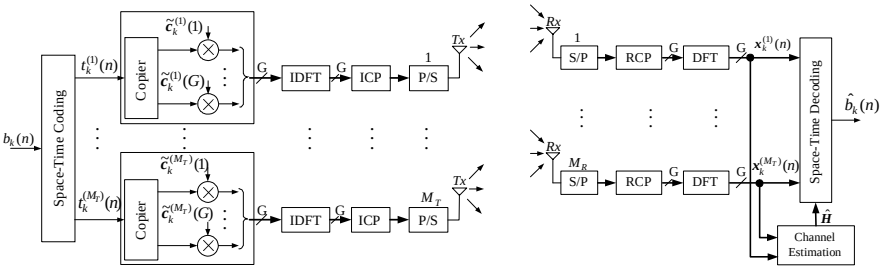


Fig. 1. Equivalent Baseband Model of MIMO MC-CDMA

The MC-CDMA scheme does the spreading spectrum (SS) operation in the frequency domain [3]. We define the m-th assigned frequency domain spreading code for the k-th user as a $G\times 1$ vector $\tilde{\boldsymbol{c}}_k^{(m)}(k=1,\cdots,K;\ \ m=1,\cdots,M_T)$, which can be written as

$$\tilde{\boldsymbol{c}}_k^{(m)}=[\tilde{c}_k^{(m)}(1)\ \ \tilde{c}_k^{(m)}(2)\ \ \cdots\ \ \tilde{c}_k^{(m)}(G)]^T \tag{1}$$

As well known, the multipath nature of the wireless channel can be described as a finite impulse response (FIR) filter, which easily results in hostile ISI. When the MC-CDMA system is over an FIR channel, an acceptable approach for combating the resultant hostile ISI is with the aid of a cyclic prefix (CP) to each transmitted data

block, where the length of CP exceeds the FIR channel memory duration. The CP can convert the time-domain linear convolution of the FIR channel to a cyclic convolution, or effectively, to a multiplicative channel distortion in the frequency domain so that the ISI can be avoided. Such multiplicative channel distortion in the frequency domain can be converted from the time-domain channel impulse response (CIR) that presents the effects of all multipaths.

For clarity, we describe the time-domain FIR CIR vector for the *k*-th user between the *p*-th (p=1,…,M_R) receive antenna and the *m*-th (m=1,…,M_T) transmit antenna as

$$\boldsymbol{h}_k^{(pm)} = [h_k^{(pm)}(0), h_k^{(pm)}(1), \cdots, h_k^{(pm)}(L-1)]^T \tag{2}$$

where L expresses the maximum common FIR-channel length of all channels between receive antennas and transmit antennas. Without loss of generality, we assume $L<G$.

The effects of above dispersive channels appear as random frequency-domain attenuations in each subcarrier. We depict the frequency-domain attenuations for the *k*-th user on all subcarrier channels between the *p*-th receive antenna and the *m*-th transmit antenna as the vector $\boldsymbol{\varsigma}_k^{(pm)}$ with dimensions $G\times 1$. Successively, we can obtain $\boldsymbol{\varsigma}_k^{(pm)}$ by performing the discrete Fourier transform (DFT) on the aforementioned time-domain FIR vector $\boldsymbol{h}_k^{(pm)}$, that is,

$$\boldsymbol{\varsigma}_k^{(pm)} \triangleq [\eta_k^{(pm)}(1) \quad \eta_k^{(pm)}(2) \quad \cdots \quad \eta_k^{(pm)}(G)]^T = \boldsymbol{F}_{DFT}(:,1:L)\boldsymbol{h}_k^{(pm)} = \boldsymbol{F}_{FRO}\boldsymbol{h}_k^{(pm)} \tag{3}$$

where the matrix $\boldsymbol{F}_{DFT}$ denotes the $G\times G$ DFT matrix with its (l, l')th element given by $(\boldsymbol{F}_{DFT})_{l,l'} = 1/\sqrt{G}\exp(-j2\pi(l-1)(l'-1)/G)$. $(l,l'=1,\ldots,G)$. The matrix $\boldsymbol{F}_{FRO}$ with dimensions $G\times L$ consists of the first L column vectors of the DFT matrix.

In multi-user MIMO MC-CDMA systems, when all K active users within one cell share all subcarriers at the same time, provided that the symbol synchronization among different users is done using timing synchronization techniques for the uplink, the frequency-domain ISI-free matrix-vector at the *p*-th receive antenna from all transmit antennas can be given by

$$\boldsymbol{x}^{(p)}(n) = \sum_{k=1}^{K}\boldsymbol{x}_k^{(p)}(n) = \sum_{k=1}^{K}\sum_{m=1}^{M_T}\mathbf{O}_k^{(m)}\boldsymbol{\varsigma}_k^{(pm)}t_k^{(m)}(n) = \sum_{k=1}^{K}\mathbf{O}_k\boldsymbol{C}_k^{(p)}\boldsymbol{t}_k(n) = \tilde{\mathbf{O}}\tilde{\boldsymbol{C}}^{(p)}\boldsymbol{t}(n) \tag{4}$$

where the matrix $\mathbf{O}_k^{(m)}$ with dimensions $G\times G$ is defined as $\mathbf{O}_k^{(m)} = diag(\tilde{\boldsymbol{c}}_k^{(m)})$; the matrix $\mathbf{O}_k$ with dimensions $G\times M_T G$ is defined as $\mathbf{O}_k = [\mathbf{O}_k^{(1)} \quad \mathbf{O}_k^{(2)}\cdots\mathbf{O}_k^{(M_T)}]$; the matrix $\tilde{\mathbf{O}}$ with dimensions $G\times KM_T G$ is defined as $\tilde{\mathbf{O}} = [\mathbf{O}_1 \quad \mathbf{O}_2\cdots\mathbf{O}_K]$; the matrix $\boldsymbol{C}_k^{(p)}$ with dimensions $GM_T\times M_T$ is the channel frequency response matrix for *k*-th user between the *p*-th receive antenna and all transmit antennas, which is defined as

$$\boldsymbol{C}_k^{(p)} = diag(\boldsymbol{\varsigma}_k^{(p1)} \quad \boldsymbol{\varsigma}_k^{(p2)}\cdots\boldsymbol{\varsigma}_k^{(pM_T)}) = \begin{bmatrix} \boldsymbol{\varsigma}_k^{(p1)} & & \\ & \ddots & \\ & & \boldsymbol{\varsigma}_k^{(pM_T)} \end{bmatrix} \tag{5}$$

$\tilde{\boldsymbol{C}}^{(p)} = \boldsymbol{I}_{\boldsymbol{K}} \otimes \boldsymbol{C}_k^{(p)}$, and $\boldsymbol{I}_K$ is a $K \times K$ identity matrix and $\otimes$ denotes the matrix Kronecker product; the vector $\boldsymbol{t}_k(n)$ defined as $\boldsymbol{t}_k(n) = [t_k^{(1)}(n), \cdots, t_k^{(M_T)}(n)]^T$, which describes the symbol block of the *k*-th user and consists of M_T symbols from M_T transmit antennas, and $\boldsymbol{t}(n) = [\boldsymbol{t}_1(n)^T, \cdots, \boldsymbol{t}_K(n)^T]^T$.

Stacking these frequency-domain ISI-free received data vectors corresponding to all M_R base station receive antennas and omitting the symbol index *n* for ease of notation, we obtain an extended ISI-free received data vector with dimensions $M_R G \times 1$, which is the numerical model for the multiuser MIMO MC-CDMA system by

$$\boldsymbol{x} = [\boldsymbol{x}^{(1)T} \quad \boldsymbol{x}^{(2)T} \quad \cdots \boldsymbol{x}^{(M_R)T}]^T = \mathbf{O}\boldsymbol{C}\boldsymbol{t} \tag{6}$$

where the matrix $\mathbf{O}$ with dimensions $M_R G \times M_R K M_T G$ is defined as $\mathbf{O} = \boldsymbol{I}_{M_R} \otimes \tilde{\mathbf{O}}$; the matrix $\boldsymbol{C}$ with dimensions $M_R K G M_T \times K M_T$ is defined as $\boldsymbol{C} = [\tilde{\boldsymbol{C}}^{(1)T} \quad \tilde{\boldsymbol{C}}^{(2)T} \cdots \tilde{\boldsymbol{C}}^{(M_R)T}]^T$.

Considering the thermal noise, we rewrite (6) as

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n} \tag{7}$$

where every entry of the thermal noise vector $\boldsymbol{n}$ with dimensions $M_R G \times 1$ is the independent identically distributed (i.i.d.) complex zero-mean Gaussian noise with variance σ_n^2.

3 Blind MUD for MIMO MC-CDMA System with ULA

Based on the aforementioned numerical model for the multi-user MIMO MC-CDMA system with the base station ULA, we present an blind MUD approach for the MIMO MC-CDMA system with the ULA.

3.1 Subspace Based Channel Identification Scheme for MIMO MC-CDMA

Now, we consider the auto-correlation matrix of $\boldsymbol{x}$ in (6). We define it as

$$\boldsymbol{R}_{\boldsymbol{xx}} = \mathrm{E}[\boldsymbol{x}\boldsymbol{x}^H] = \mathrm{E}[(\mathbf{O}\boldsymbol{C}\boldsymbol{t})(\mathbf{O}\boldsymbol{C}\boldsymbol{t})^H] = \mathbf{O}\boldsymbol{C}\boldsymbol{C}^H\mathbf{O}^H \tag{8}$$

where $\boldsymbol{R}_{\boldsymbol{xx}}$ is of dimensions $M_R G \times M_R G$. $(\bullet)^H$ denotes the Hermitian transpose.

Likewise, we describe the auto-correlation matrix of $\boldsymbol{y}$ in (7) as

$$\boldsymbol{R}_{yy} = \mathrm{E}[\boldsymbol{y}\boldsymbol{y}^H] = \mathbf{O}\boldsymbol{C}\boldsymbol{C}^H\mathbf{O}^H + \sigma_n^2 \boldsymbol{I}_{\boldsymbol{M_R G}} \tag{9}$$

If the matrix $\boldsymbol{C}$ has full column rank of KM_T and $M_R G > KM_T$ [5], $\mathbf{O}\boldsymbol{C}$ has full column rank of KM_T. Performing an eigenvalue decomposition (ED) on $\boldsymbol{R}_{\boldsymbol{yy}}$ results in two subspaces, namely the signal subspace and the noise subspace. Then, the signal subspace, $\boldsymbol{U}_S$, where signals of different users lie, is spanned by eigenvectors that corresponds to the KM_T nonzero largest eigenvalues of $\boldsymbol{R}_{yy}$, that is range{$\boldsymbol{U}_S$}=range{$\mathbf{O}\boldsymbol{C}$}. While the noise subspace defined as $\boldsymbol{U}_N = [\boldsymbol{u}_0 \quad \boldsymbol{u}_1 \cdots \boldsymbol{u}_{(M_R G - KM_T - 1)}]$ is spanned by eigenvectors associated with the remaining $(M_R G - KM_T)$ smallest eigenvalues.

Accordingly, the CIR vector can be efficiently identified other than an ambiguous complex coefficient. We further use the finite alphabet property of transmitted symbols to estimate the ambiguous complex coefficient so that the CIR vector is accurately estimated.

By exploiting the orthogonality between the signal and noise subspace, the following matrix equation holds, i.e.

$$\boldsymbol{U}_N^H \mathbf{O} \boldsymbol{C} = \boldsymbol{0}_{(M_R G - K M_T) \times K M_T} \tag{10}$$

from which we have

$$\begin{aligned} &[\boldsymbol{U}_{N1}^H \quad \boldsymbol{U}_{N2}^H \cdots \boldsymbol{U}_{NM_R}^H][\boldsymbol{I}_{M_R} \otimes \tilde{\mathbf{O}}][\tilde{\boldsymbol{C}}^{(1)T} \cdots \tilde{\boldsymbol{C}}^{(M_R)T}]^T \\ &= [\boldsymbol{U}_{N1}^H \tilde{\mathbf{O}} \quad \boldsymbol{U}_{N2}^H \tilde{\mathbf{O}} \cdots \boldsymbol{U}_{NM_R}^H \tilde{\mathbf{O}}][\tilde{\boldsymbol{C}}^{(1)T} \cdots \tilde{\boldsymbol{C}}^{(M_R)T}]^T = \boldsymbol{0}_{(M_R G - K M_T) \times K M_T} \end{aligned} \tag{11}$$

where the sub-matrix $\boldsymbol{U}_{Np}^H$ consists of M_R column vectors of the matrix $\boldsymbol{U}_N^H$ from the (p-1)M_R-th column to the pM_R-th column. We have

$$[\boldsymbol{U}_{N1}^H \mathbf{O}_k \cdots \boldsymbol{U}_{NM_R}^H \mathbf{O}_k][\boldsymbol{C}^{(1)T} \cdots \boldsymbol{C}^{(M_R)T}]^T = \boldsymbol{0}_{(M_R G - K M_T) \times K M_T} \tag{12}$$

Let $\boldsymbol{h}_k^{(m)} = [\boldsymbol{h}_k^{(1m)T} \quad \boldsymbol{h}_k^{(2m)T} \cdots \boldsymbol{h}_k^{(M_R m)T}]_{(M_R L \times 1)}^T$ denotes the channel vector for k-th user formed by the wireless FIR channels between the m-th transmit antenna and all receive antennas. From (12) we have

$$[\boldsymbol{U}_{\mathrm{N1}}^{\mathrm{H}} \mathbf{O}_k^{(m)} \cdots \boldsymbol{U}_{\mathrm{NM_R}}^{\mathrm{H}} \mathbf{O}_k^{(m)}][\boldsymbol{I}_{\mathrm{M_R}} \otimes \boldsymbol{F}_{\mathrm{FRO}}] \boldsymbol{h}_k^{(m)} = \boldsymbol{0}_{(M_R G - K M_T) \times 1} \tag{13}$$

When $M_R G - M_T K \geq M_R L$, the above linear equation set is overdeterminated. Define $\boldsymbol{Q}_k^{(m)} = [\boldsymbol{U}_{N1}^H \mathbf{O}_k^{(m)} \cdots \boldsymbol{U}_{NM_R}^H \mathbf{O}_k^{(m)}][\boldsymbol{I}_{M_R} \otimes \boldsymbol{F}_{FRO}]$. Then the equations described in (13) are equivalent to the following equations.

$$\left\| \boldsymbol{Q}_k^{(m)} \boldsymbol{h}^{(m)} \right\|^2 = \boldsymbol{h}^{(m)H} \boldsymbol{Q}_k^{(m)H} \boldsymbol{Q}_k^{(m)} \boldsymbol{h}^{(m)} = 0 \tag{14}$$

We can obtain the solution to above equations by solving the following subspace fitting problem.

$$\begin{aligned} \hat{\boldsymbol{h}}_k^{(m)} &= \arg \min_{\left\| \boldsymbol{h}_k^{(m)} \right\|^2 = 1} \sum_{k \in [1,K]} [\boldsymbol{h}_k^{(m)H} \boldsymbol{Q}_k^{(m)H} \boldsymbol{Q}_k^{(m)} \boldsymbol{h}_k^{(m)}] \\ &= \arg \min_{\left\| \boldsymbol{h}_k^{(m)} \right\|^2 = 1} \sum_{k \in [1,K]} [\boldsymbol{h}_k^{(m)H} \boldsymbol{\Theta}_k^{(m)} \boldsymbol{h}_k^{(m)}] \end{aligned} \tag{15}$$

where $\boldsymbol{\Theta}_k^{(m)} = \boldsymbol{Q}_k^{(m)\mathrm{H}} \boldsymbol{Q}_k^{(m)}$.

Obviously, we can obtain the solution of (15) by the least-squares method. In short words, the estimated channel vector $\hat{\boldsymbol{h}}_k^{(m)}$ between the m-th transmit antenna and all receive antennas is just the eigenvector of $\boldsymbol{\Theta}_k^{(m)}$ associated with its smallest eigenvalue.

By constructing $\boldsymbol{\Theta}_k^{(m)}$ (m=1, …, M_T) according to (15) for different transmit antenna, we can easily estimate the uplink channel vector in turn. Note that there is an ambiguous complex coefficient $\gamma_k^{(m)}$ between the estimated channel vector $\hat{\boldsymbol{h}}_k^{(m)}$ and the actual channel vector $\boldsymbol{h}_k^{(m)}$.

3.2 Identification of Ambiguous Complex Coefficient and Blind Multiuser Detector

Due to the uncertain complex coefficient between the estimated CIR and the original one, we further use the finite alphabet property of transmitted symbols to estimate the ambiguous complex coefficient so that the accurate CIR is obtained [6].

We substitute $\varsigma_k^{(pm)}$ (k=1~K, p=1~M_R) in (3) by $\hat{\varsigma}_k^{(pm)}$, where $\boldsymbol{h}_k^{(pm)}$ is substituted by $\gamma_k^{(m)}\hat{\boldsymbol{h}}_k^{(pm)}$,and $\gamma_k^{(m)} = |\gamma_k^{(m)}| e^{j\phi_k^{(m)}}$, $\hat{\boldsymbol{h}}_k^{(pm)} = \hat{\boldsymbol{h}}_k^{(m)}((p-1)*L+1:p*L)$. And then we obtain $\boldsymbol{C} = \hat{\boldsymbol{C}}\boldsymbol{\Gamma}$ in (6). Accordingly, (7) is rewritten as

$$\boldsymbol{y} = \boldsymbol{O}\hat{\boldsymbol{C}}\boldsymbol{\Gamma}\boldsymbol{t} + \boldsymbol{n} \tag{16}$$

where the ambiguous complex coefficient matrix $\boldsymbol{\Gamma}$ with dimensions $KM_T \times KM_T$ is described as

$$\boldsymbol{\Gamma} = \boldsymbol{I}_K \otimes diag(\gamma_k^{(1)}, \cdots, \gamma_k^{(M_T)}) = |\boldsymbol{\Gamma}| e^{j\boldsymbol{\Phi}} \tag{17}$$

where the modulus matrix $|\boldsymbol{\Gamma}|$ and the phase angle matrix $e^{j\boldsymbol{\Phi}}$ with dimensions $KM_T \times KM_T$ are defined as

$$\begin{aligned} |\boldsymbol{\Gamma}| &= \boldsymbol{I}_K \otimes diag(|\gamma_k^{(1)}|, \cdots, |\gamma_k^{(M_T)}|) \\ e^{j\boldsymbol{\Phi}} &= \boldsymbol{I}_K \otimes diag(e^{j\phi_k^{(1)}}, \cdots, e^{j\phi_k^{(M_T)}}) \end{aligned} \tag{18}$$

By defining the matrices above, we can obtain the identification of the modulus matrix and the phase angle matrix as in [6], that is

$$|\boldsymbol{\Gamma}| = \sqrt{(\boldsymbol{O}\hat{\boldsymbol{C}})^+ (\boldsymbol{R} - \sigma_n^2 \boldsymbol{I})((\boldsymbol{O}\hat{\boldsymbol{C}})^+)^H} \tag{19}$$

$$e^{j\boldsymbol{\Phi}} = \sqrt{diag(\boldsymbol{z})} \tag{20}$$

where $(\bullet)^+$ represents the Penrose-Moore pseudo-inverse; $\boldsymbol{R}$ denotes the auto-correlation matrix of the received data matrix $\boldsymbol{Y}$, which is defined as $\boldsymbol{R} = \frac{1}{L_b} \mathrm{E}[\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{H}}]$,and L_b is the number of received samples [7]; the vector $\boldsymbol{z}$ is the column-wise average of matrix $[(\boldsymbol{O}\hat{\boldsymbol{C}}|\boldsymbol{\Gamma}|)^+\boldsymbol{Y}] \cdot [(\boldsymbol{O}\hat{\boldsymbol{C}}|\boldsymbol{\Gamma}|)^+\boldsymbol{Y}]$.

Using the estimated ambiguity coefficient matrix $\boldsymbol{\Gamma}$, we obtain the equivalent spatial-temporal signature waveform matrix as $\boldsymbol{O}\hat{\boldsymbol{C}}\boldsymbol{\Gamma}$.

Once the equivalent spatial-temporal signature waveform matrix is estimated, the equivalent spatial-temporal minimum mean square error (MMSE) multi-user detector is defined as

$$\hat{\boldsymbol{t}} = ((\boldsymbol{O}\hat{\boldsymbol{C}}\boldsymbol{\Gamma})(\hat{\boldsymbol{C}}\boldsymbol{\Gamma})^H + \sigma_n^2 \boldsymbol{I})^{-1}(\hat{\boldsymbol{C}}\boldsymbol{\Gamma})\boldsymbol{y} \tag{21}$$

Where $\hat{\boldsymbol{t}}$ with dimensions $M_T K \times 1$ is the estimated vector for the transmitted symbol vector $\boldsymbol{t}$.

3.3 Brief Summary of Proposed Multi-User Detector

The proposed blind channel estimation scheme for uplink channel in MIMO MC-CDMA systems with the base station ULA can be summarized as follows:

1) Estimate the auto-correlation matrix of the received data vector $\boldsymbol{y}$ in (9) from the sampled average matrix as $\hat{\boldsymbol{R}}_{\boldsymbol{yy}} = \frac{1}{L_b}\sum_{n=1}^{L_b} \boldsymbol{y}\boldsymbol{y}^H$, where L_b depicts the number of symbol blocks;

2) Perform eigen decomposition operation on the auto-correlation matrix $\hat{\boldsymbol{R}}_{\boldsymbol{yy}}$, and construct the estimated noise subspace matrix $\hat{\boldsymbol{U}}_N^H$ according to (10);

3) Construct $\boldsymbol{\Theta}_k^{(m)}$ for each active user based on (13) and (14), and estimate the channel vector $\hat{\boldsymbol{h}}_k^{(m)}$ (k=1~K, m=1~M_T) according to (15), which is just the eigenvector associated with the minimum eigenvalue of the constructed $\boldsymbol{\Theta}_k^{(m)}$;

4) Estimate the modulus matrix $|\boldsymbol{\Gamma}|$ and the phase-angle matrix $e^{j\boldsymbol{\Phi}}$ of ambiguous complex coefficients by (19) and (20), respectively;

5) Estimate the transmitted symbol vector $\hat{\boldsymbol{t}}$ by (21).

To significantly reduce the computation complexity, we propose to estimate the CIR vector $\boldsymbol{h}_k^{(pm)}$ (p=1~M_R) for each element of the ULA concurrently for the k-th user as in [8]. Based on these estimated CIRs for all elements of the ULA, we can construct the equivalent CIR vector $\boldsymbol{h}_k^{(m)}$ for the k-th user according to (13). By exploiting these reconstructed CIRs, an equivalent multi-user detector for MIMO MC-CDMA system with the ULA can be developed as (21).

Because there is no need of a large amount of computing burden, this concurrent scheme is greatly suitable for the fast time varying wireless channels.

4 Simulation Results

Computer simulations have been conducted to demonstrate the performance of our algorithm, and a set of computer simulation results are presented to illustrate the advantages of the proposed MUD for the MIMO MC-CDMA system with ULA. The validity of our method is also illustrated. We use an MC-CDMA system with the differential binary phase shift keying (DBPSK) modulation mode in our simulations.

In simulations, we show the performance results of Monte Carlo simulation. The base station is equipped with three transmit antennas. Spreading factor G is set as 32. Hadamard multiple access codes are distributed to different users. 50 Monte-Carlo trials are performed for each simulation. The performance of our algorithm is evaluated by the root mean square error (MSE) which is defined as

$$MSE = \frac{1}{\|\hat{\boldsymbol{t}}\|}\sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}\|\hat{\boldsymbol{t}}(i)-\boldsymbol{t}\|^2}, \qquad N_t = 50 \tag{23}$$

where $\hat{\boldsymbol{t}}(i)$ is the estimate of $\boldsymbol{t}$ from the i-th trial, and $\|\bullet\|$ represents the Frobenius norm. To calculate the RMSE, it is assumed that the mean power of all users is identical.

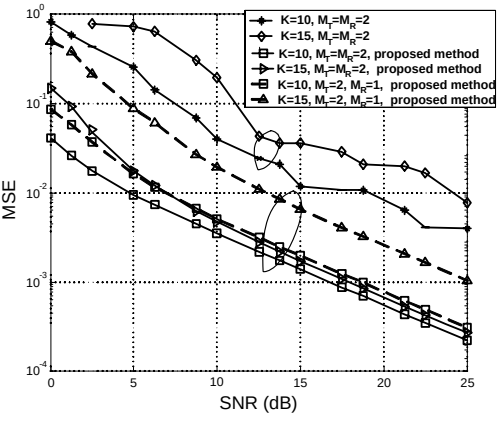


Fig. 2. The root MSE verses the SNR

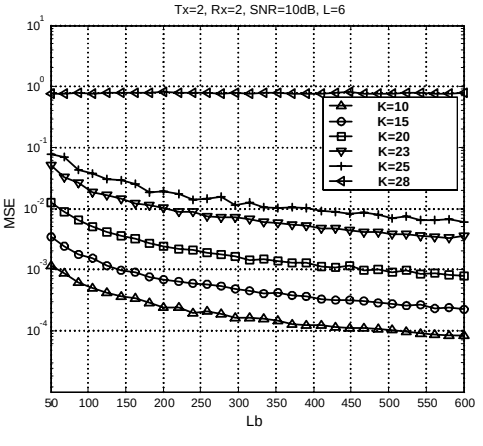


Fig. 3. The root MSE verses the L_b

In Fig.2 we present the simulation results of root MSE when the signal-to-noise ratio (SNR) varies from 0 to 25 dB, where the SNR is defined as $\mathrm{SNR} = 10\log_{10} 1/\sigma_n^2 (\mathrm{dB})$, and L=6. As it is shown in Fig.2 that the root MSE of the multi-user detector as in (21) quickly declines as the SNR increases.

The performance of the root MSE against the number of data samples L_b are displayed in Fig.3, where SNR=10dB, M_T=2, M_R=2, and L=6. According to the linear equation set (13), when $M_R G - M_T K < M_R L$, the linear equation set (13) is

underdeterminated. The uplink channel vector $\boldsymbol{h}_k^{(m)}$ (m=1,…,M_T) therefore is unable to be estimated uniquely. As shown in Fig.3, when the number of users within one cell is greater than 26, the performance of the root MSE is therefore severely degraded.

5 Conclusion

In this paper, we interpret the MIMO MC-CDMA system with the base station ULA. Then, by utilizing the known specific spreading code and the finite alphabet property of transmitted symbols, we propose an exact uplink multi-user channel estimation and an MUD scheme for MIMO MC-CDMA system over the frequency-selective fading channel. Superior to the MC-CDMA system with a single antenna at base station, our algorithm is more robust against noise and can well mitigate MAI in multi-user scenarios. Our proposed algorithm is especially significant for the larger number of antennas. Simulation results are presented for the evaluation of the proposed algorithm.

References

1. Yee N., Linnartz J-P., and Fettweis G. : Multicarrier CDMA in indoor wireless radio networks. Proc. IEEE PIMRC'93, Yokohama, Japan (1993) 109-113
2. Hara S. and Prasad R. : Overview of Multicarrier CDMA. IEEE Communication Magazine, December (1997) 126-133
3. Bingham J. A. C. : Multicarrier modulation for data transmission: An idea whose time has come, IEEE Commun. Mag., Vol. 28, May (1990) 5-14
4. Ham S. and Prasad R. : Design and performance of multicamer CDMA system in frequency-selective Rayleigh fading channels. IEEE Trons. Veh. Techno., Vol. 48, September (1999) 1584-1594
5. TONG L, XU G, KAILATH T. : A new approach to blind identification and equalization of multipath channels. Proc. Int. Conf. Commun., June (1992) 1513~1517
6. Bofeng Jiang, Xiaojun Wu, and Qinye Yin : Blind uplink channel estimation and multiuser detection for wideband CDMA. Proc. WCC-ICSP 2000, Beijing, China, August (2000) 1798-1801
7. SUN W, LI H B. : A subspace-based channel identification algorithm for forward link in space-time coded MC-CDMA systems. WCNC2002, 2002 IEEE, Vol. 1, March (2002) 17-21
8. Wu Xiaojun, Yin Qinye, and Feng Aigang : Equivalently blind time-domain channel estimation for MC-CDMA system over frequency-selective fading channels in multi-user scenario. Proc. IEEE VTC, USA (2001)

Optimal Prototype Filters for Near-Perfect-Reconstruction Cosine-Modulated Nonuniform Filter Banks with Rational Sampling Factors*

Xuemei Xie, Guangming Shi, Wei Zhong, and Xuyang Chen

School of Electronic Engineering, Xidian University, Xi'an, P.R. China
xmxie@see.xidian.edu.cn,
gmshi@xidian.edu.cn,
{wzhong, xychen}@mail.xidian.edu.cn

Abstract. In this paper, we propose a simple design method of nonuniform filter banks (NUFBs) with rational sampling factors. The analysis filters are respectively generated by cosine modulating several prototype filters with certain matching condition. By following a cosine roll-off function of the transition bands of the stretched prototype filters, the overall system is approximately power complementary and therefore possesses the near-prefect-reconstruction (NPR) property. Further, due to the use of the Parks-McClellan algorithm in the filter design, no objective function in the optimization is explicitly required and optimal prototype filters for NPR cosine-modulated NUFBs can be obtained. Design examples along with comparisons show that the resulting filter banks are of high filter quality and the proposed method is of simplicity and effectiveness.

1 Introduction

Filter banks are used in many signal processing applications. The theory of perfect reconstruction (PR) and near perfect reconstruction (NPR) filter banks with uniform frequency partition has been well established [1, 2]. However, in some applications, filter banks with flexible time-frequency decomposition are much preferred. Efficient design of nonuniform filter banks (NUFBs) is therefore preferred. On the theory and design of NUFBs, many works have been done over the years [3]-[10]. For the filter banks with rational sampling factors, detailed discussion can be found in [4]. Design methods based on modulation were reported in [5]-[9]. Recombination nonuniform cosine modulated filter banks (CMFBs) were proposed in [5], which is performed with a two-stage structure. It has a long system delay due to its recombination architecture. In [6], each filter of a NUFB is formed by merging the relevant filters of a uniform modulated filter bank, preserving the design procedure and structure of uniform CMFBs. Another nonuniform pseudo quadrature mirror filter (QMF) bank design was proposed in [8], where nonuniform filters are achieved by joining different uniform filters with transition filters. To match the different transition bandwidth, the transition filters have to be modulated from complex lowpass prototypes. Based on

* Work supported by NSF of China (Grant No. 60372047) and NSF of Shaanxi Province, China (Grant No. 2005F18).

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 245 – 253, 2006.

the cosine modulation of different linear phase prototypes, Argenti [9] proposed a new method for nonuniform QMF banks. The aliasing cancellation imposes constraints on the prototypes being dependent on each other. Among all the designs for NUFBs, most methods rely on a kind of nonlinear optimization. Since the optimization procedures are in general computationally intensive, the converging rate towards optimum solutions is rather slow. Even worse, their cost functions usually have many local minima which make the starting points choosing critical. As a result, significant human intervention is required in order to obtain an acceptable solution.

In this paper, we propose a simple design method for NPR cosine modulated NUFBs with rational sampling factors. Based on the cosine modulation of different linear phase prototypes introduced by Argenti [9], we employ the Parks-McClellan algorithm to design optimal prototype filters. Following a cosine roll-off characteristic of the transition bands of all the stretched prototype filters, the desired magnitude responses are specified as an optimization objective. Rather than resulting in a local optimal by nonlinear optimization, a global optimum solution is obtained. As a result, the overall system is approximately power complementary and hence has the NPR property. In addition, the stopband attenuation of all the analysis/synthesis filters can be significantly high. The design examples show that the performance of the results obtained by our proposed approach is much better than that in [9] with an assumption of the same distortion level. Most importantly, it also demonstrates that the proposed design is very simple.

The rest of the paper is organized as follows. Section II gives a brief introduction to the principle of NPR cosine modulated NUFBs. The design method of optimal prototype filters is described in Section III. Design examples along with comparisons are shown in Section IV. Finally, some conclusions are drawn in Section V.

2 NPR Cosine-Modulated NUFBs with Rational Sampling Factors

In this section, we will introduce the design of pseudo-QMF banks with rational sampling factors. The filter bank is generated by modulating different prototype filters. The basic idea for the design is from Argenti [9]. Here, we will briefly review the basic principle of the NUFBs along with the analysis of the elimination of the phase and significant aliasing distortions. Those will lay the foundation for our proposed method for the optimal prototype filter designs.

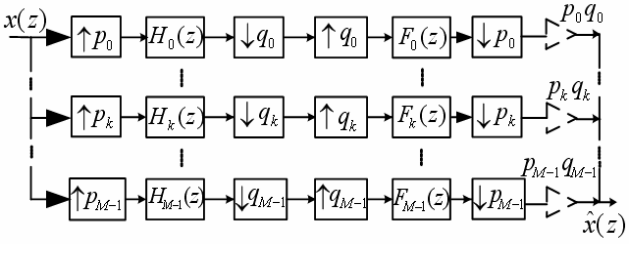


Fig. 1. General structure of NUFBs with rational sampling factors

2.1 Basic Principle of NUFBs with Rational Sampling Factors

Consider the structure of the NUFBs with rational sampling factors p_k/q_k shown in Fig. 1, where $k = 0, \cdots, M-1$. We assume that each filter is FIR with real coefficients and the sampling factors p_k and q_k are coprime. For critical sampling, they have to satisfy the condition $\sum_k p_k/q_k = 1$. The input-output relationship of this system can be expressed as

$$\begin{aligned}\hat{X}(z) &= \sum_{k=0}^{M-1}\sum_{i_k=0}^{p_k-1}\sum_{l_k=0}^{q_k-1} X\left(zW_{q_k}^{l_k p_k}\right)H_k\left(z^{1/p_k}W_{p_k}^{i_k}W_{q_k}^{l_k}\right)F_k\left(z^{1/p_k}W_{p_k}^{i_k}\right)\\ &= X(z)T(z)+\sum_{k=0}^{M-1}\sum_{l_k=1}^{q_k-1} X\left(zW_{q_k}^{l_k p_k}\right)A_{k,l_k}(z),\end{aligned} \tag{1}$$

where $W_M = e^{-j2\pi/M}$, $H_k(z)$ and $F_k(z)$ are respectively analysis and synthesis filters. The transfer function of the overall system $T(z)$ is

$$T(z) = \sum_{k=0}^{M-1}\sum_{i_k=0}^{p_k-1} H_k(z^{1/p_k}W_{p_k}^{i_k})F_k(z^{1/p_k}W_{p_k}^{i_k}), \tag{2}$$

and the aliasing components are

$$A_{k,l_k}(z) = \sum_{i_k=0}^{p_k-1} H_k(z^{1/p_k}W_{p_k}^{i_k}W_{q_k}^{l_k})F_k(z^{1/p_k}W_{p_k}^{i_k}). \tag{3}$$

The analysis filters $h_k(n)$ and synthesis filters $f_k(n)$ are assumed to be time-reversed. That is,

$$f_k(n) = h_k(N_k - n) \text{ or } F_k(z) = z^{-N_k}H_k\left(z^{-1}\right),\ 0 \le k \le M-1, \tag{4}$$

where N_k is the order of analysis filter $h_k(n)$.

Substituting (4) into (2), we get

$$T(e^{j\omega}) = \sum_{k=0}^{M-1}\sum_{i_k=0}^{p_k-1} e^{j\omega(-N_k/p_k)}W_{p_k}^{-i_k N_k}\,|H_k(e^{j\omega/p_k}W_{p_k}^{i_k})|^2 . \tag{5}$$

To eliminate the phase distortion, the following condition should be satisfied,

$$\frac{N_0}{p_0} = \frac{N_1}{p_1} = \frac{N_2}{p_2} = \cdots = \frac{N_{M-1}}{p_{M-1}} = C, \tag{6}$$

where C is a positive integer. Under this relation, $T(e^{j\omega})$ in (5) can be rewritten as

$$T\left(e^{j\omega}\right)=\sum_{k=0}^{M-1}\sum_{i_k=0}^{p_k-1}e^{-j\omega C}\left|H_k\left(e^{j\omega/p_k}W_{p_k}^{i_k}\right)\right|^2=e^{-j\omega C}\sum_{k=0}^{M-1}\sum_{i_k=0}^{p_k-1}\left|H_k\left(e^{j\omega/p_k}W_{p_k}^{i_k}\right)\right|^2. \tag{7}$$

It is clear that $T(z)$ has linear phase property and thus the phase distortion is eliminated.

2.2 Cancellation of Significant Aliasing Distortion

In [9], each analysis/synthesis filter is obtained by the cosine modulation of a lowpass prototype filter. Different prototypes may be used in different branches. Thus the modulation formulas of the *k*-th analysis and synthesis filters are respectively written as:

$$h_k(n)=2p_k(n)\cos((2l_k+1)\frac{\pi}{2q_k}(n-\frac{N_k}{2})+\theta_k), \tag{8}$$

$$f_k(n)=2p_k(n)\cos((2l_k+1)\frac{\pi}{2q_k}(n-\frac{N_k}{2})-\theta_k), \tag{9}$$

where $0\le k\le M-1$, $p_k(n)$ is the prototype filter of the *k*-th channel, l_k is the position parameter, and θ_k is the original phase. It is well known that θ_k can only be selected as $\pm\pi/4$ for the significant aliasing cancellation in the uniform CMFBs. From (9), it can be found that the time-reverse relation of analysis and synthesis filters is still maintained. With phase distortion of the system being eliminated, we turn to the cancellation of aliasing distortion.

To begin with, we introduce two concepts: high side and low side transition bands of a filter. For the high side, we mean the transition band in the higher frequency region; while for the low side, we refer to the one in the lower frequency region.

For the cancellation of the aliasing error, certain requirements must be met in the following cases,

$$\text{for the } \textit{high-high} \text{ case, } e^{\mp j2\theta_k}+e^{\pm j2\theta_{k+1}}=0, \tag{10}$$

$$\text{for the } \textit{high-low} \text{ case, } e^{\mp j2\theta_k}+e^{\mp j2\theta_{k+1}}=0, \tag{11}$$

$$\text{for the } \textit{low-high} \text{ case, } e^{\pm j2\theta_k}+e^{\pm j2\theta_{k+1}}=0, \tag{12}$$

$$\text{for the } \textit{low-low} \text{ case, } e^{\pm j2\theta_k}+e^{\mp j2\theta_{k+1}}=0. \tag{13}$$

The detailed derivation can be found in [9]. To minimize the aliasing distortion further, it is necessary to give a lemma as follows:

Lemma: Let $T_{l(k)}$ and $T_{r(k)}$ denote the width of the left and right transition bands of $H_k(z)$, partitioning the lower and higher frequency components of input signal respectively. To minimize the aliasing distortion further, we should set $T_{r(k)}p_k=T_{l(k+1)}p_{k+1}$, $0\le k\le M-2$.

3 Design of Optimal Prototype Filters

Since the reconstructed signal does not suffer from phase distortion, and the significant aliasing error is also reduced, the only consideration for us is the amplitude distortion. In this section, we will deal with this remaining problem and design the optimal prototype filters.

Due to the advantages of the Parks-McClellan algorithm, it is employed in our design of optimal prototype filters. Thus, the filter performance is an equiripple optimum approximation and the magnitude response of the filter is completely specified and used as the design target. The Parks-McClellan algorithm for filter designs is also very popular in constructing filter banks with a linear phase property. Interested readers please refer to [10].

In the nonuniform CMFBs, all the analysis and synthesis filters are the shifted and stretched versions of different prototype filters. By imposing some constraints on the transition bands of all the stretched prototypes, an ideal filter magnitude response can be achieved. To make $\left|T(e^{j\omega})\right|$ be approximately flat, it is sufficient to impose the following constraint on the prototype filters $P_k(e^{j\omega})$,

$$\left|\sum_{i_k=0}^{p_k-1} P_k(e^{j\omega/p_k} W_{p_k}^{i_k})\right|^2 + \left|\sum_{i_{k+1}=0}^{p_{k+1}-1} P_{k+1}(e^{j(\omega-\varphi_{k+1})/p_{k+1}} W_{p_{k+1}}^{i_{k+1}})\right|^2 = 1, \text{ for } 0 < \omega < \varphi_{k+1}, \tag{14}$$

where $\varphi_{k+1} = \frac{p_k \pi}{2q_k} + \frac{p_{k+1}\pi}{2q_{k+1}}$, $0 \le k \le M-2$. We notice that this constraint in the transition band would be met if we force the response of the right transition band of the stretched version $\sum_{i_k=0}^{p_k-1} P_k(e^{j\omega/p_k} W_{p_k}^{i_k})$ in the interval $\omega \in [0,\pi]$ to follow a cosine function and that of the left transition band of the stretched and shifted version $\sum_{i_{k+1}=0}^{p_{k+1}-1} P_{k+1}(e^{j(\omega-\varphi_{k+1})/p_{k+1}} W_{p_{k+1}}^{i_{k+1}})$ to follow a sine function such that $\cos^2\alpha + \sin^2\alpha = 1$, where $0 \le \alpha \le \pi/2$. This constraint is fundamental to our design.

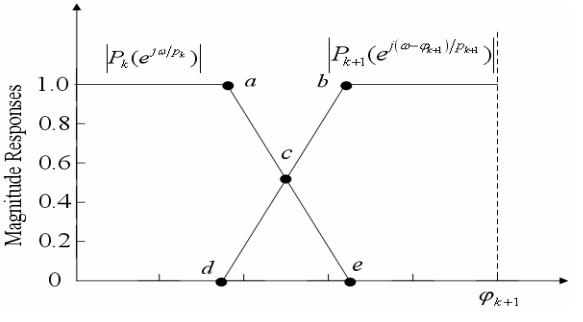


Fig. 2. The magnitude responses of the stretched and shifted versions of adjacent two prototype filters in the region $\omega \in [0, \varphi_{k+1}]$

Fig. 2 illustrates the desired responses of $\sum_{i_k=0}^{p_k-1} P_k(e^{j\omega/p_k} W_{p_k}^{i_k})$ and $\sum_{i_{k+1}=0}^{p_{k+1}-1} P_{k+1}(e^{j(\omega-\varphi_{k+1})/p_{k+1}} W_{p_{k+1}}^{i_{k+1}})$ in the region $\omega \in [0,\pi]$, denoted by $\left|P_k(e^{j\omega/p_k})\right|$ and $\left|P_{k+1}(e^{j(\omega-\varphi_{k+1})/p_{k+1}})\right|$ respectively. Points *a* and *e* indicate respectively the passband and stopband edges of $\left|P_k(e^{j\omega/p_k})\right|$, denoted by $\omega_p^{(k)}$ and $\omega_s^{(k)}$. While points *b* and *d* associated with $\left|P_{k+1}(e^{j(\omega-\varphi_{k+1})/p_{k+1}})\right|$ are the counterparts of points *a* and *e*. Point *c* is the intersection of the magnitude responses of the two filters. Clearly, if the value at this point is $1/\sqrt{2}$, then the sum of the square amplitude of the two filters at this point will be equal to unity. In addition, it should be expected that (a) $\left|P_k(e^{j\omega/p_k})\right|$ and $\left|P_{k+1}(e^{j(\omega-\varphi_{k+1})/p_{k+1}})\right|$ are symmetric with respect to point *c*, (b) point *c* corresponds to $\omega_c^{(k)} = p_k\pi/2q_k$ which is the cutoff frequency of $P_k(e^{j\omega/p_k})$, and (c) the passband edge frequency $\omega_p^{(k)}$ of $\left|P_k(e^{j\omega/p_k})\right|$ should be equal to the stopband edge frequency $\omega_s^{(k+1)}$ of $\left|P_{k+1}(e^{j(\omega-\varphi_{k+1})/p_{k+1}})\right|$, which should be held inversely. Now we are ready to characterize their magnitude responses in the transition band. Denote $A(\omega)$ as the response of $\left|P_k(e^{j\omega/p_k})\right|$ in the right transition band and $B(\omega)$ that of $\left|P_{k+1}(e^{j(\omega-\varphi_{k+1})/p_{k+1}})\right|$ in the left. To meet all the constraints and requirements described above, we must set $A(\omega)$ and $B(\omega)$ to follow cosine and sine functions in their transition bands respectively. Finally, it can be found that the transition bands of all the stretched prototype filters are similar with each other under this constraint.

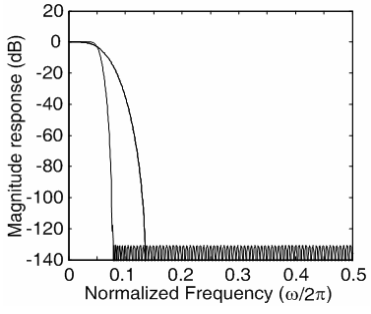
4 Design Examples

In this section, we will give two examples and compare the results with those in [9] to show the simplicity and effectiveness of the proposed method.

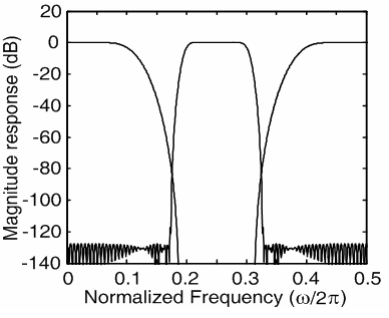
The first example is a 3-channel NUFB with sampling rate $[1/5, 3/5, 1/5]$. Here, the sampling factors $p_0 = p_2 = 1$, $p_1 = 3$, and $q_0 = q_1 = q_2 = 5$. Due to the fact that $p_0 = p_2$, only two prototype filters need to be designed. To eliminate the phase distortion, Eq. (6) should be satisfied. Here we set $C = 55$. Thus the order of analysis/synthesis filters is $[55, 165, 55]$. The couplings of the aliasing components that must be considered are of the type high-high and low-low between the branches 0-1 and 1-2, respectively, so we select $\theta = [\pi/4, \pi/4, \pi/4]$. By employing the Parks-McClellan algorithm and using Eqs. (8), (9) and (14), we get the desired filter bank. Fig. 3 shows the magnitude responses of prototype filters and analysis filters, as well

as the amplitude and aliasing distortions. Their maximum values denoted by E_{pp} and E_a are $E_{pp} = 6.463\times10^{-3}$ and $E_a = 7.452\times10^{-4}$, respectively. And the stopband attenuation is 117.5 dB.

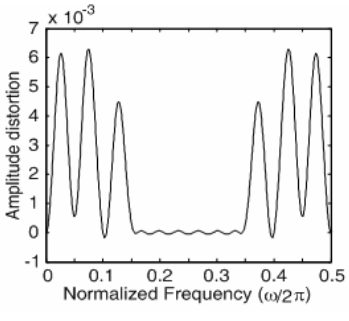
While in example of [9], under the same design parameters, the maximum amplitude and aliasing distortion are about $E_{pp} = 1.6\times10^{-3}$, $E_a = 2.5\times10^{-4}$, and the stopband attenuation is only approximately 70 dB.



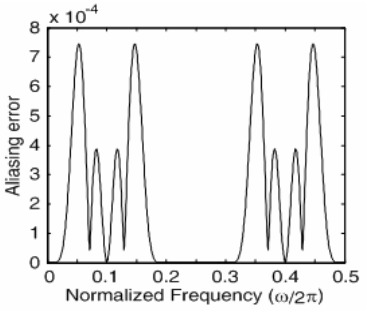
(a) Magnitude responses of prototype filters



(b) Magnitude responses of analysis filters



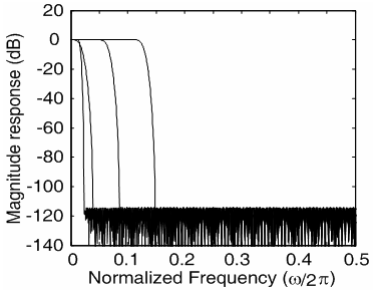
(c) Amplitude distortion



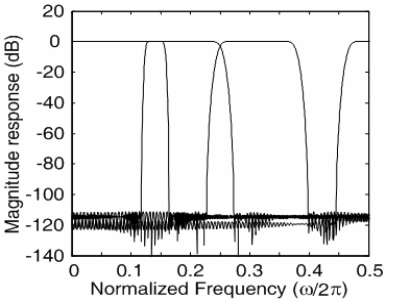
(d) Aliasing error

Fig. 3. A 3-channel NPR cosine modulated NUFB with sampling rate [1/5, 3/5, 1/5]

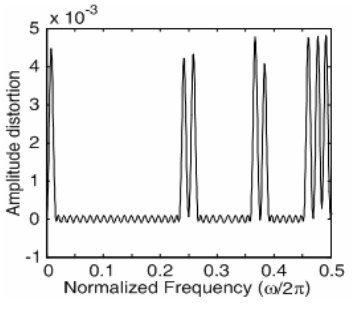
The second example is a 4-channel nonuniform CMFB with sampling rate $[1/2, 1/4, 3/16, 1/16]$. Here we choose $C = 169$, and the order of analysis/synthesis filters is $[169, 169, 507, 169]$. All the aliasing couplings are of the type high-low, thus we get $\theta = [\pi/4, -\pi/4, \pi/4, -\pi/4]$. Also by following the same design procedures, we obtain the desired result. The magnitude responses of prototype filters and analysis filters are shown in Fig. 4, along with the amplitude and aliasing distortion. Their maximum values are $E_{pp} = 4.903\times10^{-3}$ and $E_a = 1.826\times10^{-3}$, respectively. And the stopband attenuation is 111.2 dB.



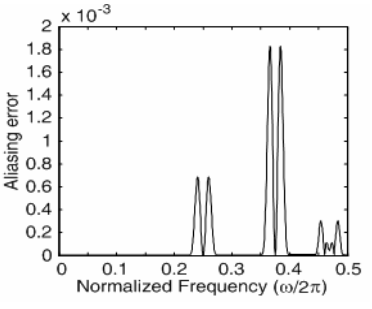
(a) Magnitude responses of prototype filters



(b) Magnitude responses of analysis filters



(c) Amplitude distortion



(d) Aliasing error

Fig. 4. A 4-channel NPR cosine modulated NUFB with sampling rate [1/2, 1/4, 3/16, 1/16]

Again it is interesting to compare the result of our second example with the one with the same decimation factors given in [9]. Under the same design specifications such as filter length and stopband cutoff frequency, the amplitude and aliasing distortion given in [9] are about 3.8×10^{-3} and 3.2×10^{-3} respectively, which are almost the same as ours. However, the stopband attenuation by using our approach is nearly 60 dB higher than that in [9] (111.2 dB versus 50 dB). More importantly, the analysis and synthesis filters in our design are capable of achieving the comparable performance without nonlinear optimization.

5 Conclusions

In this paper, a simple approach is proposed for designing optimal prototype filters for NPR nonuniform CMFBs with rational sampling factors. By employing the Parks-McClellan algorithm, we force the transition bands of all the stretched prototype filters to follow a cosine function so that the overall system possesses a NPR property. The design examples have shown that the comparable performance can be achieved in terms of significantly higher stopband attenuation. Most importantly, it is very efficient and effective without nonlinear optimization.

References

1. Vaidyanathan, P.P.: Multirate Systems and Filter Banks. Prentice Hall, Englewood Cliffs, New Jersey (1992)
2. Truong, T.Q.: Near-Prefect-Reconstruction Pseudo-QMF Banks. IEEE Trans. on Signal Processing, Vol. 42, (1994) 65–75
3. Nayebi, K., Barnwell, T.P., Smith, M.J.T.: Nonuniform Filter Banks: A Reconstruction and Design Theory. IEEE Trans. on Signal Processing, Vol. 41, No. 3, (1993) 1114–1127
4. Kovacevic, J., Vetterli, M.: Perfect Reconstruction Filter Banks with Rational Sampling Factors. IEEE Trans. on Signal Processing, Vol. 42, No. 6, (1993) 2047–2066
5. Xie, X.M., Chan, S.C., Yuk, T.I.: On the Design of a Class of PR Nonuniform Cosine Modulated Filter Banks with Flexible Rational Sampling. IEEE Trans. on Circuits and Systems I, Vol. 52, No. 9, (2005) 1965–1981
6. Lee, Jeong-Jin, Lee, Byeong Gi: A Design of Nonuniform Cosine Modulated Filter Banks. IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing, Vol. 42, No. 11, (1995) 732–737
7. Li, J., Nguyen, T.Q., Tantaratana, S.: A Simple Design Method for Near-Perfect-Reconstruction Nonuniform Filter Banks. IEEE Trans. on Signal Processing, Vol. 45, (1997) 2105–2109
8. Princen, J.: The Design of Nonuniform Modulated Filter Banks. IEEE Trans. on Signal Processing, Vol. 43, No. 11, (1995) 2550–2560
9. Argenti, F., Brogelli, B., Re, E.D.: Design of Pseudo-QMF Banks with Rational Sampling Factors Using Several Prototype Filters. IEEE Trans. on Signal Processing, Vol. 46, No. 6, (1998) 1709–1715
10. Xie, X.M., Chan, S.C.: Theory and Design of a Class of Nonuniform Filter Banks with Linear Phase Filters. IEICE Trans. on Information and System, Vol. E88_D, No. 7, (2005) 1445–1452

XRMCCP: A XCP Framework Based Reliable Multicast Transport Protocol

Guang Lu[1], YongChao Wang[2], and MiaoLiang Zhu[1]

[1] Institute of Artificial Intelligence, Zhejiang University, Hangzhou, 310027, China
{luguang, zhum}@zju.edu.cn
[2] Center of Network &Information, Zhejiang University, Hangzhou, 310027, China
Yongchaowang@zju.edu.cn

Abstract. This paper introduces a XCP framework based reliable multicast transport protocol XRMCCP (XCP framework based Reliable Multicast Congestion controlled protocol) that has been designed to be simple, scalable and reliable for high bandwidth-delay product (BDP) networks. We observe that TCP Friendly Throughput Equation (TFTE), which many current multicast transport protocols use to estimate available rates of multicast sessions, has limits of performance in high BDP networks. XRMCCP generalizes XCP framework to support one-to-many transport congestion control. The scalability issue is addressed with an exponential timers scheme that is also used to estimate the number of receivers involved in the communication. The paper presents a number of simulation results on its performance. Besides, some design choices are evaluated.

1 Introduction

Multicast technique was thought as a very efficient transmission method in one-to-many application such as content distribution, VOD, interactive simulation, coordinated virtual reality environment. For multicast to be successful, it is crucial that multicast transport protocol has effective congestion control scheme to respond to congestion in the bottleneck links. A larger number of multicast congestion control schemes, such as TFMCC [1], PGMCC [2], ORMCC [12] and RCCMP [8] etc., have been proposed. In order to be TCP-friendly, all of abovementioned schemes require to monitor the behavior of each receiver and to measure many parameters such as packet loss rate (P) and Round Trip Time (RTT), which are filled into TFTE to estimate available data sending rate. Those multicast congestion control mechanisms perform well in low BDP networks.

However, TCP friendly throughput equation based multicast congestion control mechanism has limits in high BDP networks. Firstly, TFTE cannot fully characterize the rate changing of TCP. TFTE is the TCP response function describing the steady-state sending rate of TCP. TCP has another important state, slow start. There are many short-term flows such as HTTP and telnet, etc., which only behavior in the slow start phase.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 254 – 263, 2006.

Second, multicast congestion control mechanisms imitate the mechanism of current TCP versions, while papers [4], [5], [6], [7], [10] show that current TCP versions perform poorly in high bandwidth-delay product networks. With development of network, links with 10Gbps or even more capability will be deployed in the Internet. For example, the capability of backbone links of the next generation Internet of China reaches 20Gbps. Both TCP's AIMD and TFTE prevent flows from achieving high throughput, and decrease the utilization of high bandwidth links. More and more applications, which apply multicast to deliver large volume of data in Local Area Networks (LAN), will be deployed in Metropolitan Area Networks (MAN) and Wide Area Networks (WAN). Not only unicast congestion controlled flows take the responsibility to efficiently use resource in high BDP networks, but also multicast congestion control mechanism should be made some modification to contribute to do that.

Third, in order to generate status report, receivers of TFTE based multicast congestion controlled session should allocate buffer to store information of received packets, while receivers should allocate more buffer size to cope with increasing number of packets in high BDP networks. If the size of buffer isn't assigned appropriate value, it is prone to come out wrong status report of network. We give some result of simulation to further explore the limits of TFTE in the next section.

The recently developed Explicit Control Protocol (XCP) [10] is a congestion control system well suited for high-BDP networks. It delivers the highest possible application performance over a range of network infrastructure, including extremely high-speed and high-delay links not well served by TCP. It achieves maximum link utilization and reduces wasted bandwidth due to packet loss. It separates the efficiency and fairness policies of congestion control, enabling routers to quickly make use of available bandwidth while conservatively managing the allocation of bandwidth to flows. It is built on the principle of carrying per-flow congestion state in packets. And its packets each carry a small congestion header through which the sender requests a desired throughput. Routers compute a fair per-flow bandwidth allocation without maintaining any per-flow state. The sender thus learns of the bottleneck router's allocation in a single round trip. It effectively decouples congestion control from bandwidth-allocation policy. The router has two separate controllers: the Efficiency Controller for preventing congestion and maintaining high utilization and the Fairness Controller for splitting the aggregate bandwidth fairly among connections sharing the link.

This paper introduces a XCP framework based reliable multicast transport protocol XRMCCP (XCP framework based Reliable Multicast Congestion controlled protocol) that has been designed to be simple, scalable and reliable for high bandwidth-delay product (BDP) networks. We observe that TCP Friendly Throughput Equation (TFTE), which many current multicast transport protocols use to estimate available rates of multicast sessions, has limits of performance in high BDP networks. XRMCCP generalizes XCP framework to support one-to-many transport congestion control. The scalability issue is addressed with an exponential timers scheme that is also used to estimate the number of receivers involved in the communication. The paper presents a number of simulation results on its performance. Besides, some design choices are evaluated.

2 Evaluation of TFTE

In order to compete fairly with TCP, many multicast congestion control protocols and TFRC use the TFTE, which roughly describes TCP's sending rate as a function of the loss event rate, round-trip time, and packet size. A loss rate is calculated by constructing a loss history and identifying loss events. These events are then converted to a loss event rate. Smoothed RTT, loss event, and various other values are then used as inputs into TFTE [3]:

$$X = \frac{s}{R\sqrt{\frac{2bp}{3}} + t_{RTO}\left(3\sqrt{\frac{3bp}{8}}\right) p\left(1+32p^2\right)} \tag{1}$$

Which calculates a TCP-compatible transmission rate X (bytes/sec) where s is the packet size (bytes). R is the round trip time (sec), p is the loss rate, tRTO is the TCP retransmission timeout (sec), and b is the number of packets acknowledged by a single TCP acknowledgement. Updates in bandwidth availability are made at a frequency of once every RTT.

However, the available bandwidth estimated by the control equation is not friendly with TCP flows, especially in high-BDP networks. Firstly, the TFTE cannot fully characterize the rate changing of TCP. The TFTE is the TCP response function describing the steady-state sending rate of TCP. TCP has another important state, slow start. There are many short-term flows such as: HTTP, telnet, etc. only in the slow start phase. Second, the input parameter p is assigned with an inaccurate estimated value. If the value of p is higher than the real value, the estimated throughput is lower than the deserved throughput, else the flow become more aggressive than TCP flows. The loss rate estimation scheme affects the performance of TCP friendly flows. Third, in high BDP networks, flows often maintain larger congestion window, so receivers should allocate larger buffer to store packets information. If the size of buffer isn't assigned appropriate value, it is prone to come out wrong status report of network. Besides, Paper [9] listed some limits of TFTE.

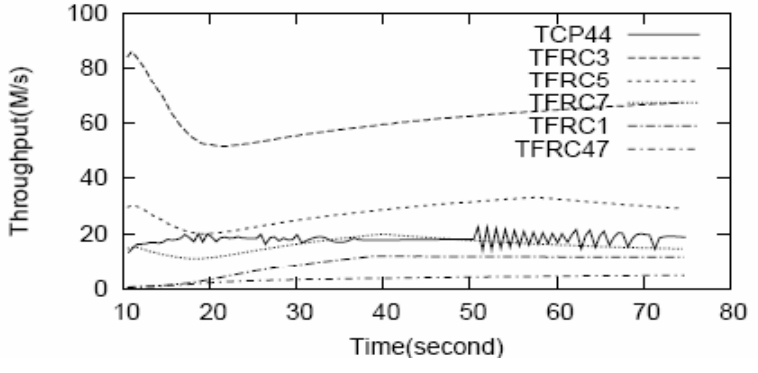


Fig. 1. The throughput of TCP flows vs. the throughput of TFTE-based TFRC flows

We carry out simulations in ns-2 [11]. The result is shown in the Figure 1. The simulations use a dumbbell network topology with a common bottleneck link between all senders and receivers. Each sender/receiver is connected to the bottleneck link by a high bandwidth access link. The propagation delays of the access link are varied in the range from 2ms to 10ms. The propagation delay in the bottleneck is 40ms. The routers use FIFO queuing with RED queue management. The maximum packet size is 1048. There are 32 TCP flows and 32 TFRC flows running in the simulations.

In the Figure 1 throughput of TCP flows and TFRC flows are shown. We observed that TFTE based TFRC flows are not friendly with TCP flows, and the utilization of the bottleneck is very low, about 73.5%.

3 Description of XRMCCP Protocol

XRMCCP is a XCP framework based reliable multicast congestion control protocol for high-speed networks. XRMCCP is a window based single rate reliable multicast transport congestion control protocol. Senders, intermediate routers and receivers all are based on XCP framework to select a appropriate transmission rate, especially in the high bandwidth-delay product network, to fairly share available bandwidth with other traffic flows and to increase bandwidth utilization while not cause congestion incur. XRMCCP uses congestion header, which is on top of TCP/UDP layer, to convey congestion information such as sender's current congestion window W, receiver representative ID, Round trip time between sender and receiver representative, delta throughput. Senders maintain congestion window W, receiver representative ID and round trip time. After senders initiate congestion header with these variables, Senders send packets to following router. Intermediate Routers receive packets arriving them and adopt efficient controller and fair controller to update delta throughput value. Only representative need feedback congestion information by ACK. Details of receiver's behavior and receiver representative selection is presented in following subsection.

3.1 Representative Selection

The main goal of representative selection is to select the slowest receiver among the group members. At the beginning of a session, the sender sends an initial hello packet (IHP) to all the group members. As soon as receiver receives the IHP, the sender sends an IHPACK to the sender as quickly as possible. Once the sender gets the first of IHPACK from one of receives, the receiver that sent the IHPACK is selected as transient representative. This selected receivers is one of the closest to the sender. To find out the slowest receiver, the sender will change the representative when a NAK from one of receivers is received. To avoid frequent changes, selected representative lasts at least two times round trip time (RTT). The representative feedbacks ACK to the sender, then, the sender will adjust its congestion window according to the equation: W = W + Delta. To avoid stall of convergence when remaining available bandwidth approach zero, XCP router allocates available bandwidth equally to every packet across itself. Addition to this, XCP router's fair controller allocates 10 percent of arrival rate as bandwidth shuffle in order to achieve max-min fairness among

flows. Many experiments show that XCP protocol has very small convergence time. So the sender will find out the slowest receiver after several RTT. RTT in the congestion header doesn't affect correctness of delta throughput calculated by routers that are not in the path from the sender to representative.

During the rest of the communication, a representative change is caused by every one or both of two following cases:

$$\begin{aligned} &\text{Case1: packet loss found by one receiver;} \\ &\text{Case2: } W_l + \sum Delta < W_{cur} \end{aligned} \tag{2}$$

Where W_l is the sender's congestion window at the end of last round. When a receiver sends a ACK, it includes its deserved congestion window and highest received sequence number. The sender compares them with the representative's ones, and if they are worse enough, the new receiver will be selected as the slowest receiver. To avoid excessive representative changes, the selection mechanism is biased towards the current representative. Besides, each representative should stay a minimum period on this state.

Another cause to look for a new representative is that the current one has left the session. If four retransmissions of a packet are sent and there is no reply, the sender has to request IHP to receivers in order to select the slowest receiver.

3.2 Flow Control

In XRMCCP, a closed loop control is run between the sender and the representative. A window-based congestion control scheme that emulates XCP congestion control is used. So, the representative behavior is equivalent to XCP's receiver in a point-to-point connection. When a packet is received, the slowest receiver executes the following algorithm:

1. If the sequence number is the expected, the representative sends an ACK packet to sender.
2. If the sequence number is lower than the expected, the received packet is a repair one that replies to a sent NAK. This packet is not acknowledged.
3. If the sequence number is higher than the expected, an ACK packet with the expected sequence number is sent immediately to acknowledge the last received packet and to notify the sender that some packets are lost.

This algorithm is executed at the representative and it is a general XCP algorithm. At the sender, a standard XCP algorithm is also executed, but packets instead of bytes are used to control transmission window. The sender algorithm can be summarized as follows:

1. The window sized is initiated depending on packet size.
2. Packets are sent until the window is full. The retransmission timer is started.
3. ACK packets are expected and are used to update RTT. (This time is estimated between the sender and the representative and is calculated in seconds).
    1) If the sequence number is the expected, every ACK with congestion header, the window size is update by : W+delta.

2) If the sequence number is not the expected and three duplicate ACKs are received, then congestion is detected. The window size is decreased, the last non- acknowledged packet is resent and the retransmission timer is reset.

If ACK is not received and the retransmission timer expires, the retransmission timer is doubled, and the window size is set to 1.

3.3 Receiver's Behavior

Receivers know current the sender's current congestion window and round trip time between the sender and the representative. Receivers can estimate the sender's sending rate R = W/T. Receives integrate delta throughput in every T and remember the sender's congestion window when one round starts. Two round later, receivers calculate their estimated congestion window and compare with the sender's current congestion window. If estimated congestion window is less than the sender's current congestion window, the receiver can send an ACK to the sender. If it is really the slowest one among the group, it will be selected as the representative.

3.4 Error Control

XRMCCP achieves reliability by means of NAK mechanism. How does a receiver distinguish a loss packet that representative has already received. Since representative can send NAK to the sender when some packets are lost, while other receivers cannot intermediately send NAK to the sender to inform the sender to retransmit them. XRMCCP need a mechanism to assure other receivers can inform the sender when it's necessary. The representative is allowed to multicast ACK to the group members. In case of that representative has acknowledged a packet while a receiver doesn't yet, if this receiver's delta throughput is smaller than the representative's this receiver is allowed to send a NAK intermediately to the sender. This receiver is selected as new representative. If not, the receiver should wait for random time sending NAK. During the waiting time, the packet arrive the receiver, the receiver cancels sending NAK. In other case of that receivers and the sender don't receive ACK from representative, the sender will retransmit the loss packets automatically.

4 Simulations

We have implemented XRMCCP on ns-2 and conducted a simulation study to validate the proposed design in terms of intra protocol fairness, inter protocol friendliness and representative selection.

We consider network topologies with a single bottleneck link (dumbbell-like), multiple bottleneck links (Figure 4). Each sender/receiver is connected to the bottleneck link by a high bandwidth access link. All senders, receivers and intermediate router nodes are XCP capable. The propagation delays of the access link are varied to obtain different RTT between each sender/receiver pair. The propagation delay in the bottleneck is 20ms in short RTT scenario and 80ms in long RTT scenario respectively.

4.1 Intra Protocol Fairness

Intra protocol fairness has been analyzed by running multiple instances of the protocol through the same bottleneck link. The bottleneck has a capacity of 500 Mbps and a router buffer size of BDP/size packets, where size is packet size. In this simulation, packet size is 1009. There are n multicast connections, and n is in the range from 1 to 16. The propagation delays between the senders and the router are (10+n)ms. Each multicast connection has three receivers with a propagation delay from the router to the receivers of 10, 20 and 30ms respectively. The senders start each 10s.

In Figure 2 the fairness index [13] is shown to different number of sessions. Fairness index measures the fair division of bandwidth. A value near to one means that network resources are fairly shared. The results show that the instances share bandwidth in a fair way. When there are 16 multicast sessions running at the same time, fairness index of XRMCCP is 0.99, while fairness index of RCCMP of ref [8] is 0.94. Although we change the link delay or add random losses the intra XRMCCP fairness is achieved.

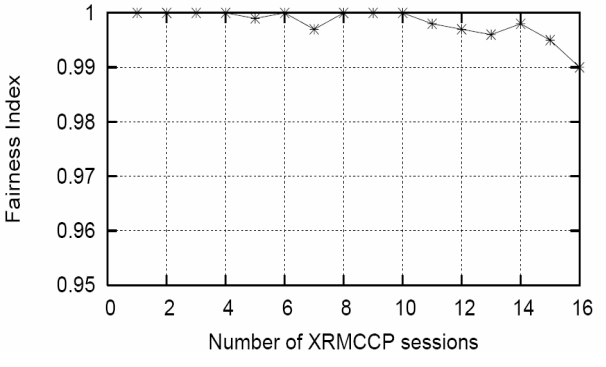


Fig. 2. Intra protocol fairness: Fairness Index of multiple XRMCCP instances

4.2 Inter Protocol Fairness

XRMCCP tries to have a behavior as similar to XCP as possible avoiding starvation in XCP connections. For this reason, the flow and congestion control mechanisms are similar to XCP.

Inter protocol friendliness has been studied by running multiple instances of XCP and XRMCCP. The bottleneck link capacity is 50*n Mbps and router buffers of n*15 packets, where n is the number of XCP sessions. The propagation delays from the router to the receivers are equal to m*10ms, where m is the number of receivers, and the propagation delay from the sources to the router is (10+s)ms, where s is the number of senders. Figure 3(a) shows the fairness index to different experiments. Figure 3(b) shows the utilization of bottleneck to different experiments. We see that XRMCCP and XCP flows fairly share the bandwidth resource. The utilization of bottleneck is very high.

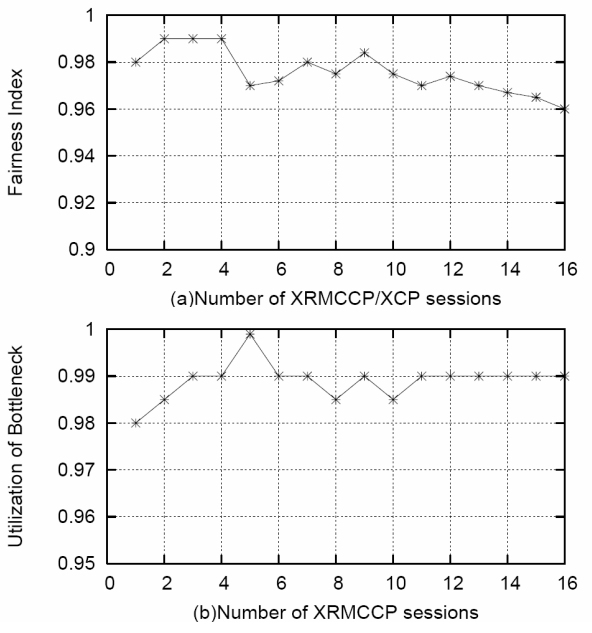


Fig. 3. (a). Inter protocol fairness: Fairness Index of XRMCCP and XCP sessions; (b) Utilization of the bottleneck; X label means n XRMCCP sessions and n XCP sessions

4.3 Representative Selection

We have run several simulations to evaluate the behavior of the representative process. The experiments involved a single sender with multiple receivers connected through independent bottlenecks. Figure 5 shows a simulation result based on the scenario proposed on Figure 4, involving one XRMCCP session with two receivers and one XCP session. The links L1 and L2 have a capacity of 500Mbps and 400Mbps

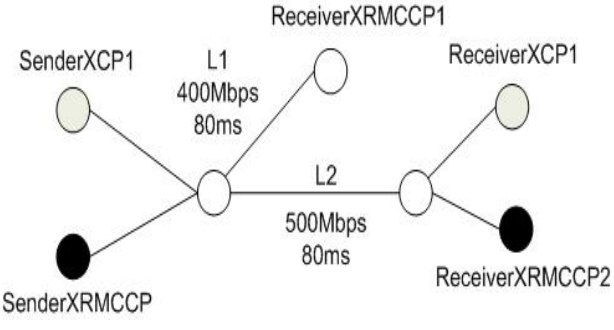


Fig. 4. Topology: One sender XRMCCP with two receivers and one XCP connection. XRMCCP starts at 2s and XCP1 at 62s.

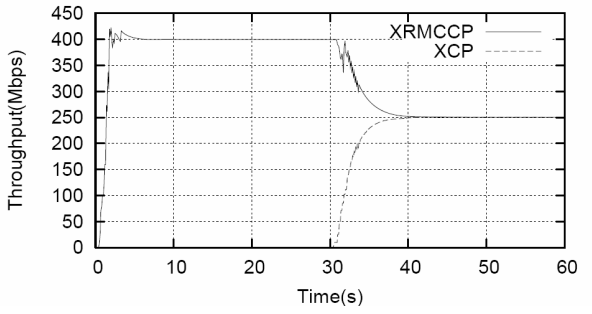


Fig. 5. Representative selection: throughput of XRMCCP and XCP sessions. At the beginning XRMCCP1 is the representative, a representative change take places at 32s and XRMCCP2 is the new representative.

respectively. Both links have a propagation delay of 80ms and buffer sizes of 100 packets. The expected behavior, which it can be observe in the graph, is that at the beginning of the session the XRMCCP takes the full 400Mbps and XRMCCP1 is the representative. When XCP1 starts the throughput is reduced to 250Mbps and a representative change takes place. The representative changes achieve that XRMCCP flow is adapted to the available bandwidth.

5 Conclusion

XRMCCP generalizes XCP framework to support one-to-many transport congestion control for high-speed networks. XRMCCP is a window based single rate reliable multicast transport congestion control protocol. As far as I know, it's the first time to adopt XCP framework in multicast transport.

Simulations show that fairness between XCP and XRMCCP connections and between XRMCCP connections are fulfilled. Besides, XRMCCP has short convergence time and can be helpful to increase the utilization of bottleneck in High-BDP networks.

As a future work, we modify XRMCCP to support multiple rate multicast transport congestion control.

References

[1] J.Widmer and M. handley, "Extending Equation-based Congestion Control to Multicast Applications", proc ACM SIGCOMM 2001, San Diego, Augest 2001.

[2] L. Rizzo. "PGMCC: A TCP-friendly single-rate multicast congestion control scheme". ACM SIGCOMM 2000. Stockholm. Augest 2000.

[3] S. Floyd, M. Handley, J. Padhye, and J. Widmer. "TCP friendly rate control (TFRC): Protocol specification". RFC 3448, January 2003.

[4] L. Xu, K. Harfoush, and I. Rhee. "Binary Increase Congestion Control (BIC) for Fast Long-Distance Networks". INFOCOM'04, March 2004.

[5] I. Rhee and L. Xu. “CUBIC: A New TCP-Friendly High-Speed TCP Variant”. PFLDNet’05, February 2005.
[6] C. Jin, D. Wei, and S. Low. “FAST TCP: Motivation, Architecture, Algorithms, Performance”. INFOCOM’04, March 2004.
[7] S. Floyd. “HighSpeed TCP for Large Congestion Windows”. IETF RFC 3649, December 2003.
[8] J. M. Kirschberg, M. S. Delgado, and S. S. Ribes. “RCCMP: A TCP-friendly Reliable Multicast Transport Protocol”. Proceedings of the 10th IEEE Symposium on Computers and Communications (ISCC 2005).
[9] Injong Rhee and Lisong Xu. “Limitations of Equation-based Congestion Control”. SIGCOMM’05 Augest 2005.
[10] D.Katabi, M. Handley, and C. Rohrs. “Congestion Control for High Bandwidth-Delay Product Networks”. SIGCOMM’02, Auguest 2002.
[11] The network simulator ns-2. http://www.isi.edu/nsnam/ns.
[12] Jiang Li, Shivkumar Kalyanaraman. “ORMCC: A Simple and Effective Single-Rate Multicast Congestion control Scheme”,submitted July 2002.
[13] D-M Chiu, R. Jain. “Analysis of the Increase and Decrease Algorithm hms for Congestion Avoidance in Computer Networks”, Computer Networks and ISDN Systems. 1989. vol. 17. pp. 1-15.

Small-World Optimization Algorithm for Function Optimization

Haifeng Du, Xiaodong Wu, and Jian Zhuang

Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System
Xi'an Jiaotong University, Xi'an 710049, China
Haifengdu@mail.xjtu.edu.cn

Abstract. Inspired by the mechanism of small-world phenomenon, some small-world optimization operators, mainly including the local short-range searching operator and random long-range searching operator, are constructed in this paper. And a new optimization algorithm, Small-World Optimization Algorithm (SWOA) is explored. Compared with the corresponding Genetic Algorithms (GAs), the simulation experiment results of some complex functions optimization indicate that SWOA can enhance the diversity of the population, avoid the prematurity and GA deceptive problem to some extent, and have the high convergence speed. SWOA is shown to be an effective strategy to solve complex tasks.

1 Introduction

The small-world phenomenon has its roots in experiments performed by the social psychologist Stanley Milgram in the 1960s to trace out short paths through the social networks of the United States [1]. The results indicate that any two individuals in the social network are all linked by short chains of acquaintances (popularly known as "six degrees of separation"). After Duncan Watts and Steve Strogatzs' paper - "Collective Dynamics of Small-World Networks" published in Nature in 1998 [2], the small-world phenomenon has been introduced to network theory by some computer scientists. And it's with great expectations to become an important theory related to computer science after systematology, information theory and cybernetics. The small-world phenomenon is an interdisciplinary problem involving with sociology, mathematics and computation science, which has become hotspots in the theoretical discussion of Artificial Intelligence (AI) and Complexity Theory. It has been successfully used in internet control [3], AIDS diffusion forecast [4] and dynamical research of biological protein networks [5], yet little discussion about small-world phenomenon has been found in the field of optimization algorithm.

If the precision is set, optimization space can be considered as a grid space (network space) constructed by the variables. We can take the optimization as a process that information transmits from candidate solution to optimal solution in search space (networks). In this paper, therefore, a new optimal search algorithm is constructed which makes use of the effective information transmission mechanism inspired by the small-world phenomenon. We proceed as follows: Section 2 provides

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 264 – 273, 2006.

a description of the general theories about small-world phenomenon and introduces the optimal search process from the point view of the small-world principle. Based on the local shortcuts search operator and random long-range search operator of small-world, section 3 discusses a kind of Small World Optimization Algorithm (SWOA). The experiments and the corresponding analyses bout SWOA are described in Sections 4. Finally, Section 5 states some conclusions.

2 Small-World Networks and Optimal Process

Roughly speaking, the small-world phenomenon reveals a most effective mode of information transmission in the movement of a lot of complex networks (such as social network), which are high clustering subnest including "local contacts" nodes and some random long-range shortcuts that conduce to produce shortcuts to increase efficiency of information transmission. The small-world phenomenon still has no accurate definition. Ordinarily, if the mean distance L between two nodes of the networks increases logarithmically with the number of network nodes N (namely $L \propto \ln N$), it can be considered that the network has the small-world phenomenon.

The small-world network of Watts and Strogatz is an intermediate between regular and random networks [2], namely, it combined with a subnet that has local clustering shortcuts and some random long-range shortcuts. Kleinberg points out that the small-world phenomenon is an efficiency question about routing algorithm, where local knowledge suffices to find effective paths to get to destination [1].

Without loss of generality, we consider the following global optimization problem P:

$$\text{maximize} \quad f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{d} \le \boldsymbol{x} \le \boldsymbol{u} \tag{1}$$

Where, $\boldsymbol{x} = \{x_1, x_2, \cdots, x_m\} \in \Re^m$ is a variable vector, $\boldsymbol{d} = \{d_1, d_2, \cdots, d_m\}$ and $\boldsymbol{u} = \{u_1, u_2, \cdots, u_m\}$ define the feasible solution space, namely, $x_i \in [d_i, u_i]$ $i = 1,2,\cdots,m$, $f(\boldsymbol{x})$ is the objective function.

When m=2, x_1-x_2 constructs the search space as figure 1 shown, in which x_1,x_2 get its value only at the circle. Let T be the optimal solution and B be the initial candidate solution, thus the resolving procedure of optimal question P is the search from B to T, which can be considered as the information transmission process from B to T, as the broken line shown in figure 1. The optimal search networks that we were talking about above is very similar to "the planar lattice small-world networks with many random contact"[2] constructed by Watts and Strogatz. If we consider the small-world phenomenon in search process, which is to say that we introduce global random search (random long-range contacts) into local search (local shortcuts) and emphasize the global result induced by local action, an optimal search algorithm with small-world phenomenon will be constructed. We call this kind of algorithms as Small world Optimization Algorithm (SWOA).

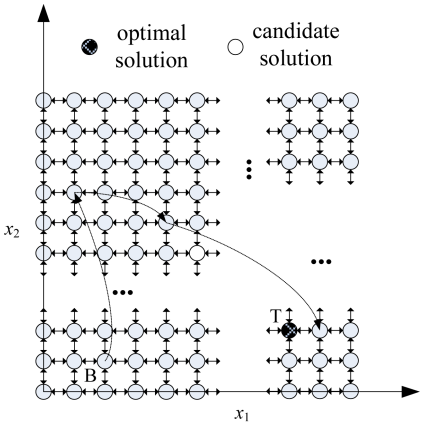


Fig. 1. Node Lattice of Two-dimensional Space Search Question

3 Small-World Operator and the Algorithm

3.1 Basic Definition

Considering the efficiency of algorithm and Stanley Milgram's design about sociological examination [1], the point which participates in search or information transmission should not be a simple one but a set of candidate solutions. We denote these transmission node populations as set $\boldsymbol{S}$. Furthermore, $\boldsymbol{s} = s_1 s_2 \cdots s_l$, $\boldsymbol{s} \in \boldsymbol{S}$, is the node coding of the variable $\boldsymbol{x}$, described by $\boldsymbol{s} = \hbar(\boldsymbol{x})$, and $\boldsymbol{x}$ is called the decoding of node $\boldsymbol{s}$, described as $\boldsymbol{x} = \hbar^{-1}(\boldsymbol{s})$. Let τ be the number of the possible values of a_i. For binary coding and decimal coding, there are $\tau = 2$ and $\tau = 10$ accordingly. Generally, antibody bit string is divided into *m* parts, and the length of each part is l_i, where $l = \sum_{i=1}^{m} l_i$, and each part is denoted as $x_i \in [d_i, u_i]$ $i = 1,2,\cdots,m$. Especially, for binary coding, we use the decoding method as follow:

$$x_i = d_i + \frac{u_i - d_i}{2^{l_i} - 1}\left(\sum_{j=1}^{l_i} s_j 2^{j-1} \right) \tag{2}$$

In this paper, we adopt 0-1 binary encoding.
And small-world search space ($\boldsymbol{s} \in \boldsymbol{I}$) is:

$$\boldsymbol{I}^n = \left\{ \boldsymbol{S} : \boldsymbol{S} = (\boldsymbol{s}_1, \boldsymbol{s}_2 \cdots \boldsymbol{s}_n),\ \ \boldsymbol{s}_k \in \boldsymbol{I},\ \ 1 \le k \le n \right\} \tag{3}$$

Where the positive integer n is the size of node population. The node population $\boldsymbol{S} = \{\boldsymbol{s}_1, \boldsymbol{s}_2 \cdots \boldsymbol{s}_n\}$, which is an n-dimension group of node $\boldsymbol{s}$, is a point in the node group space $\boldsymbol{I}^n$

The global optimal solutions for set of problem P is defined as the following

$$\boldsymbol{T} \equiv \left\{\boldsymbol{s} \in \boldsymbol{I} : f(e^{-1}(\boldsymbol{s})) = f^* \equiv \min(f(\boldsymbol{x}) : \boldsymbol{d} \le \boldsymbol{x} \le \boldsymbol{u})\right\} \tag{4}$$

Where f^* is the optimal value of objective function f. For the node population $\boldsymbol{S}$, $\vartheta(\boldsymbol{S}) \equiv |\boldsymbol{S} \cap \boldsymbol{T}|$ denotes the number of optimal solutions in antibody population $\boldsymbol{S}$.

Define the distance between $\boldsymbol{s}_i, \boldsymbol{s}_j \in \boldsymbol{S}$ is $d(\boldsymbol{s}_i, \boldsymbol{s}_j) = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$, where $\|\bullet\|$ denotes an arbitrary norm. For binary coding, we use Hamming distance; while for decimal coding we generally use Euclidean distance.

The ℓ neighborhood set of the node $\boldsymbol{s}_i$ can be defined as

$$\varsigma^{\ell}(\boldsymbol{s}_i) = \left\{\boldsymbol{s}_j \middle| 0 < \|\boldsymbol{s}_i - \boldsymbol{s}_j\| \le \ell, \quad \boldsymbol{s}_j \in \boldsymbol{S}\right\} \tag{5}$$

And $|\varsigma^{\ell}(\boldsymbol{s}_i)|$ is the number of elements in $\varsigma^{\ell}(\boldsymbol{s}_i)$, $\overline{\varsigma^{\ell}(\boldsymbol{s}_i)}$ is the non- ℓ neighborhood set of $\boldsymbol{s}_i$. Obviously, there is

$$\overline{\varsigma^{\ell}(\boldsymbol{s}_i)} = \left\{\boldsymbol{s}_j \middle| \ell < \|\boldsymbol{s}_i - \boldsymbol{s}_j\|, \quad \boldsymbol{s}_j \in \boldsymbol{S}\right\} \tag{6}$$

Furthermore, if $\boldsymbol{s}_i \notin \boldsymbol{T}$ and $\vartheta(\varsigma^{\ell}(\boldsymbol{s}_i)) \equiv |\varsigma^{\ell}(\boldsymbol{s}_i) \cap \boldsymbol{T}| = 0$, there is $\vartheta\left(\overline{\varsigma^{\ell}(\boldsymbol{s}_i)}\right) \equiv \left|\overline{\varsigma^{\ell}(\boldsymbol{s}_i)} \cap \boldsymbol{T}\right| \ge 1$.

3.2 Local Shortcuts Search Operator

The main action of local shortcuts search operator $\boldsymbol{\Psi}$ is transmitting the information from node $s_i(k)$ to the corresponding node $s_i(k+1)$ which is the nearest node to the goal set T in $\varsigma^{\ell}(\boldsymbol{s}_i(k))$ with minor ℓ. We denote this process with $\boldsymbol{s}_i(k+1) \leftarrow \Psi(\boldsymbol{s}_i(k))$. In practice, $|\varsigma^{\ell}(\boldsymbol{s}_i(k))|$ is very big. We often randomly select $n_l(k) < |\varsigma^{\ell}(\boldsymbol{s}_i(k))|$ nodes from $\varsigma^{\ell}(\boldsymbol{s}_i(k))$ to construct a temporary local transmission network to realize local search. For example: 20 bits binary encoding node $\boldsymbol{s}_i(k)$, $|\varsigma^1(\boldsymbol{s}_i(k))| = 19$ but $|\varsigma^2(\boldsymbol{s}_i(k))| = |\varsigma^1(\boldsymbol{s}_i(k))| + C_{20}^2$. In detail, we construct the following local shortcuts search operator $\boldsymbol{\Psi}$ for binary encoding.

$\boldsymbol{s}_i'(k) \leftarrow \boldsymbol{s}_i(k) \times 0$ get a temporary zero bit string which has the same construction as $\boldsymbol{s}_i(k)$ .And $\oplus, \neg, \otimes$ corresponding to bit-oriented adding、minus and multiply.

Local Shortcuts Search Operator Ψ

given $n_l(k) < \left|\varsigma^{\ell}(\boldsymbol{s}_i)\right|$; $\boldsymbol{s}_i(k+1) \leftarrow \boldsymbol{s}_i(k)$;
begin
$j \leftarrow 0$;
repeat
$\boldsymbol{s}_i'(k) \leftarrow \boldsymbol{s}_i(k) \times 0$;
Stochastic selecting $r(0 < r \le \ell)$ bits from $\boldsymbol{s}_i'(k)$ and change them into 1;
$\boldsymbol{s}_i'(k) \leftarrow \boldsymbol{s}_i(k) \neg 2 \times (\boldsymbol{s}_i(k) \otimes \boldsymbol{s}_i'(k)) \oplus \boldsymbol{s}_i'(k)$;
if $f\left(e^{-1}(\boldsymbol{s}_i'(k))\right) < f\left(e^{-1}(\boldsymbol{s}_i(k+1))\right)$
$\boldsymbol{s}_i(k+1) \leftarrow \boldsymbol{s}_i'(k)$;
end if
$j \leftarrow j+1$;
until $j = n_l(k)$
end.

Fig. 2. Flowchart of Local Link Search Operator

3.3 Random Long-Range Search Operator

Essentially, a point in $\overline{\varsigma^{\ell}(\boldsymbol{s}_i)}$ is stochastically selected as the information transmission object point, $\boldsymbol{s}_i'(k)$ of $\boldsymbol{s}_i(k)$, by random long-range search operator $\boldsymbol{\Gamma}$ according to the preset probability distribution, which can be described as $\boldsymbol{S}'(k) \leftarrow \boldsymbol{\Gamma}(\boldsymbol{S}(k))$ (ℓ gets a bigger value, namely $\|\boldsymbol{s}_i - \boldsymbol{s}_i'(k)\|$ is bigger). Considering the convenience and availability, the Inverse Operation in evolutionary computation is used to construct the operator $\boldsymbol{\Gamma}$. The random long-range search operator $\boldsymbol{\Gamma}$ in this paper is described as the figure 3.

Where rand(0-1) generates uniform distributed random number whose values fall between 0 and 1. $\boldsymbol{s}_i'(k)\big|_{\mu}^{\nu}$ means overturning positive sequence of the characters' from bit μ to bit ν ,for example:

$$\boldsymbol{s}_i'(k) \quad \underset{\mu}{1}\ 0\ 1\ 0\ 1\ 1\ 0\ 0\ 1\ 1\ 0\ \underset{\nu}{0} \xrightarrow{\boldsymbol{s}_i'(k)\big|_{\mu}^{\nu}} 1\ 1\ 0\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ 0$$

p_l in the operator $\boldsymbol{\Gamma}$ is corresponding to the inversion probability of inversion operation in evolutionary algorithm except that a limit for the distance between μ and ν is set.

Random Long-range Search Operator Γ

Given global long-range probability p_l and ℓ;

begin

$i \leftarrow 0$;

Repeat

$s_i'(k) \leftarrow s_i(k)$;

$p \leftarrow \text{rand}(0-1)$;

if $p_l < p$ and $s_i(k)$ is not the optimal solution in current set of node,

Stochastic generate two integer μ 、 ν ,where $1 \le \mu < \nu \le l$ and $|\mu - \nu| > \ell$;

$s_i(k) \leftarrow s_i'(k)|_\mu^\nu$;

end if

$i \leftarrow i+1$;

until $i = n$

end.

Fig. 3. Flowchart of Random Long-range Search Operator

3.4 Small World Optimal Algorithm

Based on the local shortcuts operator Ψ and random long-range operator above, the step of realizing SWOA is shown in figure 4.

The stop conditions are defined as either the restricted iterative number or the time when the solutions are not improved at successively iterations, or both of the two. Combining the enactment iterative times with hunting condition, here the algorithm is halted as the following criterion:

$$\left| f^* - f^{best} \right| < \varepsilon \tag{7}$$

Where f^* is the global optimum of f, f^{best} is the current best function value. If $0 < \left| f^* \right| < 1$, the following equation holds:

$$\left| f^* - f^{best} \right| < \varepsilon \left| f^* \right| \tag{8}$$

Compared with the evolutionary algorithm, the SWOA does not adopt crossover operator or global selection operation; instead, local selecting is applied so that local information transmission in small world phenomenon is emphasized. Instead of simple point search strategy, the algorithm adopts population as in the evolutionary algorithm, which is consistent with socialist investigative experiment [1].

According to the analysis of the two-dimensional grid network just like figure 2, Jon Kleinberg proved that the expected delivery time of the decentralized algorithm is at most $\alpha(\log N)^2$ [6]. Inspired by Jon Kleinberg's analyzing and thinking of the

Small World Optimal Algorithm (SWOA)

Initialization: Enactment the size of node population n, inversion probability p_l, ℓ and other parameters. initialize the node population $\boldsymbol{S}(0)=\{\boldsymbol{s}_1(0),\boldsymbol{s}_2(0),\cdots,\boldsymbol{s}_n(0)\}$;

$k \leftarrow 0$;

begin

while terminative condition doesn't satisfy;

$\boldsymbol{S}'(k) \leftarrow \boldsymbol{S}(k)$;

$\boldsymbol{S}'(k) \leftarrow \boldsymbol{\Gamma}(\boldsymbol{S}'(k))$;

$i \leftarrow 0$

Repeat

$\boldsymbol{s}_i'(k+1) \leftarrow \boldsymbol{\Psi}(\boldsymbol{s}_i'(k))$, $\boldsymbol{s}_i'(k) \in \boldsymbol{S}'(k)$;

if $f\left(e^{-1}(\boldsymbol{s}_i'(k+1))\right) < f\left(e^{-1}(\boldsymbol{s}_i(k))\right)$

$\boldsymbol{s}_i(k+1) \leftarrow \boldsymbol{s}_i'(k+1)$;

else

$\boldsymbol{s}_i(k+1) \leftarrow \boldsymbol{s}'(k)$;

end if

$i \leftarrow i+1$;

until $I = n$

$k \leftarrow k+1$;

end while

end.

Fig. 4. Flowchart of SWOA

characteristic of SWOA, we can roughly estimate SWOA's computation complexity as $\alpha(l \log \tau)^2$.

4 Simulation Experiment

4.1 Test Function and Conditions

Seven functions adopted in our simulation experiment are shown as formulas from (9) to (15). These functions' space shapes are so complex that their optimal question can not be better solved by using general algorithm.

$$\max f_1(x,y) = 1 + x \times \sin(4\pi x) - y \times \sin(4\pi y + \pi) + \frac{\sin(6\sqrt{x^2+y^2})}{6\sqrt{x^2+y^2}+10^{-15}} \quad x, y \in [-1,1],\ f^* = 2.118 \tag{9}$$

$$\max f_2(x,y) = 1 + x \times \sin(4\pi x) - y \times \sin(4\pi y + \pi) \quad x, y \in [-1,1],\ f^* = 2.055 \tag{10}$$

$$\max f_3(x,y) = \left(\frac{b}{a+\left(x^2+y^2\right)}\right)^2 + \left(x^2+y^2\right)^2 \quad x, y \in [-5.12,5.12],\ f^* = 3600 \text{ where } a=0.05\ b=3 \tag{11}$$

$$\min f_4(x,y)=\left\{\sum_{i=1}^{5} i\cos[(i+1)x+i]\right\}\times\left\{\sum_{i=1}^{5} i\cos[(i+1)y+i]\right\} \quad x,y\in[-10,10],\ f^*=-186.73 \tag{12}$$

$$\min f_5(x,y)=\left(4-2.1x^2+x^{\frac{4}{3}}\right)x^2+xy+\left(-4+4y^2\right)y^2 \quad x,y\in[-5.12,5.12]\ \ f^*=-1.031628 \tag{13}$$

$$\max f_6(x,y)=-\left(x^2+y^2\right)^{0.25}\left(\sin^2 50\left(x^2+y^2\right)^{0.1}+1.0\right) \quad x,y\in[-5.12,5.12]\ f^*=0 \tag{14}$$

$$\min f_7(x,y)=20+x^2-10\cos(2\pi x)+y^2-10\cos(2\pi y) \quad x,y\in[-5.12,5.12]\ \ f^*=0 \tag{15}$$

All numerical experiments in this paper are implemented on PC of PIV2.4G、512MRAM by using Matlab language.

4.2 Experiment Result and Analysis

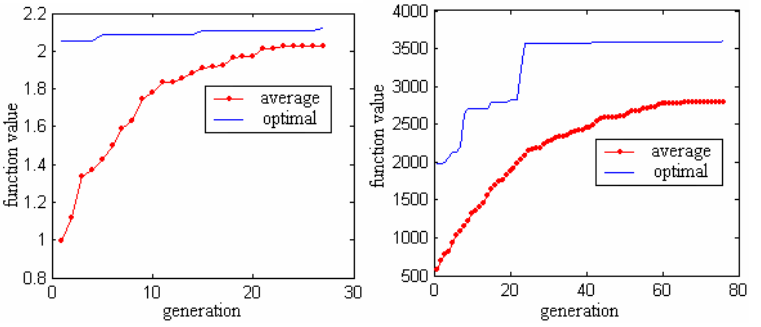
The performance comparison of SWOA and a GA (MGA) is shown in table 1. MGA is a modified genetic algorithm which adopts new selection operator and population update strategy [7]. The population size of MGA is 30 and the encoding length is 80, maximal generation is 1500, crossover probability is 0.7, mutation probability is 0.0017, Generation gap (GGAP, namely how many new individuals are created) is 0.9. In order to keep the approximate computational complexity with MGA, the population size $\boldsymbol{S}$ of SWOA is30 and $n_l(k)$ is 4, p_l is 0.3, ℓ is 1, node's encoding length is also 80.The statistical results in table 1 are obtained from 50 simulations.

Table 1. Performance comparison of MGA and SWOA

Contrast item		f_1	f_2	f_3	f_4	f_5	f_6	f_7
N_{best}	MGA	21	50	45	19	49	8	49
	SWOA	50	50	50	50	50	50	50
N_{max}	MGA	74	26	1382	706	200	1031	1005
	SWOA	113	29	1136	166	49	396	88
N_{min}	MGA	2	1	45	20	1	161	19
	SWOA	5	2	32	7	7	43	23
N_{mean}	MGA	14.48	6.24	387.9	78.95	22.43	593.	363.0
	SWOA	31.92	8.56	115.7	59.84	25.9	116.3	45.92
N_{std}	MGA	15.57	5.24	308.6	153.68	27.43	339.0	266.3
	SWOA	22.02	5.07	166.1	34.84	11.75	54.50	14.84
ε		1.0×10^{-2}	1.0×10^{-2}	1.0×10^{-1}	1.0×10^{-2}	1.0×10^{-2}	1.0×10^{-2}	1.0×10^{-2}

Where N_{bes} denote the time of reaching the optimal value, N_{max}, N_{min}, N_{mean}, and N_{std} denote the maximum, minimum , average and standard deviation of the evaluation number reaching the optimal value respectively. N_{max}, N_{min}, N_{mean}, and N_{std} in table 1 are only the statistic results without immersing into the local optimal value, because MGA can break away from the local optimal triumphantly only for function

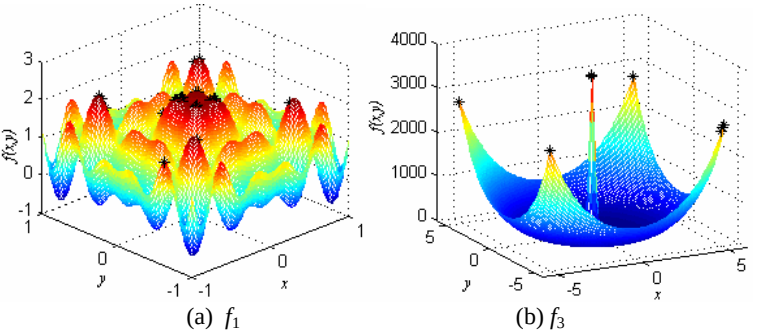
f_2 in 50 simulations. For f_2, the performance of SWOA is equivalent with MGA. For other functions, SWOA has more strong ability to break away from the local optimum than MGA; and the convergent speed of ADCSA is generally faster than MGA. N_{std} of SWOA is less than that of MGA universally, which indicate that SWOA has better robustness and stability than MGA. f_3 is a function which is used in Evolutionary Computation to test algorithm's ability to overcome deceptive problem. The optimal result shows that SWOA overcomes the deceptive problem to a certain extent. The relatively changing of the objective function denotes the algorithm's diversity essentially. For functions f_1 and f_3, the change of objective function in one experiment is shown in fig 5. It can be seen that SWOA keeps the population diversities better.



(a) f_1 (b) f_3

Fig. 5. The changing of the function value for SWOA

Figure 6 gives one of the optimization solution distributed in function space for f_1 and f_3, where "*" denotes the final solution.



(a) f_1 (b) f_3

Fig. 6. The optimized result of SWOA

Obviously, SWOA has searched a lot of local optimal solution at the same time. The corresponding evolutionary computation, such as MGA, can also search the optimal solution but the distribution of the solution is centralized, furthermore the

individual of population may become completely the same. Accordingly, SWOA has better diversity than MGA.

5 Conclusions and Prospecting

Based on the view of small-world phenomenon, the text gives a new description of optimal search process. It considers that the search process is essentially an information transmission process in search space networks. Different from six degrees of separation phenomenon experiment, we know less or even no knowledge about the behavior of object point so that the question will be more complex. Making use of theories of small-world phenomenon, the paper constructs local shortcuts search operator and random long-range search operator; furthermore it constructs a Small World Optimization Algorithm used for optimization. The experiment of some typical function optimal questions shows that the SWOA has following advantages: avoiding the problem of getting into local minimum; increasing the diversity of population; improving convergence rate and overcoming deceptive problem to a certain extent.

Main work of the paper is concentrating on constructing a new optimal algorithm based on the Small-World Phenomenon mechanism, at the same time it exploits a new field in application of small-world phenomenon. Kleinberg found that a subtle variant of the Watts–Strogatz network will in fact support efficient search: Rather than adding the long-range shortcuts uniformly at random, add links between nodes of this network with a probability that decays like the d^{th} power of their distance (in d dimensions)[1]. Therefore how to improve the searching efficiency of small world algorithm and the interrelated theoretical analysis are foci of our further research.

Acknowledgements

The support of the National Natural Science Foundation of China (NSFC) (No. 50505034) is gratefully acknowledged.

References

1. Kleinberg, J.: The Small-World Phenomenon and Decentralized Search. SIAM News, 37(3) (2004) 1–2
2. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature, 393(4) (1998) 440–442
3. Albert, R., Barabasi A.L.: Statistical mechanics of complex networks. Rev. Mod Phys, 74 (2002) 47–97
4. Liljeros, F., Edling C.R., Amaral L.A.N., et al.: The web of human sexual contacts. Nature, 411(6840) (2001) 907–908
5. Jeong, H., Tombor, B., Albert, R., et al.: The large-scale organization of metabolic networks. Nature, 407(6804)(2001) 651–654
6. Kleinberg, J.: The Small-World Phenomenon: An Algorithmic Perspective. Cornell Computer Science Technical Report, (1999) 99–1776
7. Chipperfield, A., Fonseca, C., Pohlheim, H., et al.: Genetic Algorithm TOOLBOX. http://www.shef.ac.uk/cgi-bin/cgiwrap/gaipp/gatbx-download

A Two-Dimension Chaotic Sequence Generating Method and Its Application for Image Segmentation

Xue-Feng Zhang[1,2] and Jiu-Lun Fan[2]

[1] Department of Electronic Engineering, Xidian University, Xi'an, Shaanxi, 710071, P.R. China
[2] Department of Information and Control, Xi'an Institute of Post and Telecommunications, Xi'an, Shaanxi, 710061 P.R. China
zhangxuefeng3@163.com

Abstract. Chaotic optimization is a new optimization technique. For image segmentation, conventional chaotic sequence is not fit to two-dimension gray histogram because it is proportional distributing in $[0,1]\times[0,1]$. In order to generate a chaotic sequence can be used to the optimization processing of image segmentation method in two-dimension gray histogram, we propose an chaotic sequence generating method based on Arnold chaotic system and Bézier curve generating algorithm. Simulation results show that the generated sequence is pseudorandom. The most important characteristic of this chaotic sequence is that its distribution is approximately inside a disc whose center is $(0.5, 0.5)$, this characteristic indicates that the sequence is superior to the Arnold chaotic sequence in image segmenting. Based on the extended chaotic sequence generating method, we study the two-dimension Otsu's image segmentation method using chaotic optimization. Simulation results show that the method using the extended chaotic sequence has better segmentation effect and lower computation time than the existed two-dimension Otsu's method.

1 Introduction

Image segmentation is the fundamental and first step for image analysis and understanding. Image thresholding selection method[1] is a main and applause segmentation method. Among the thresholding selection methods, Otsu's method is an attractive one due to its simplicity and effective[1,2]. But in practice, Otsu's method in one-dimension gray histogram could fail to images with noises. In order to improve the segmentation effect, Liu and Li[3] extended Otsu's method to two-dimension gray histogram, their method not only uses the pixels' gray histogram but also uses the pixels' neighborhood's information. In their method, the hypothesis that "the sum of the target and the background's probability is approximate to 1" was used and a close expression can be obtained. Considering the computation time for Otsu's method with two-dimension histogram is large, recursion algorithm was given [4]. It had found that the hypothesis is not established in many cases. Therefore, it is valued to study the two-dimension Otsu's method which not use the hypothesis.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 274 – 283, 2006.

For this case, because we could not get a close expression, so we face large computation amount. In this paper, we will use two-dimension chaotic optimization to process this problem.

Chaotic systems are nonlinear dynamic behavior, they are pseudorandom and sensitivity to the initialize conditions. Because chaotic systems have good properties, chaotic systems are widely used in communications, optimization, control and image processing et.al[5-8]. For image segmentation, Xiu et.al [9] used Logistic chaotic sequence to optimize one-dimension entropy thresholding method and better segmentation result was obtained. But it is unfeasible to directly extend the one-dimension chaotic optimization method to two-dimension situation, because the generated two-dimension chaotic sequence is proportional distributing in $[0,1]\times[0,1]$. It should be pointed out that for two-dimension gray histogram, the pixels inside the target class and the background classes' distribution are distributed near the diagonal from $(1,1)$ to (L,L), so the conventional chaotic sequences is not fit to image segmentation in the case of two-dimension gray histogram, the problem is how to generate a two-dimension sequence that is pseudorandom and the values of x and y in the sequence are approximately equality. In order to solve this problem, the key work is to construct a two-dimension chaotic sequence has the mentioned properties.

Section 2 presents the chaotic sequence generating method based on chaotic system and Bézier curve generating algorithm, Section 3 presents a two-dimension Otsu's image segmentation algorithm based on chaotic optimization(without using the hypothesis of "the sum of the target and the background's probability is approximate to 1") and also analysis the effect and the efficiency of the method by experimentations. The conclusion is given in Section 4.

2 Chaotic Sequence Generating Method

2.1 Bézier Curve and Its Properties

In 1971, Bézier presented a generating curve method based on Bernstein function and control points[10], this method is widely used in computer aided geometric design(CAGD) because of its good properties. The method is described as follows:

Suppose $P_i=(x_i,y_i,z_i)$, $i=0,1,\cdots,n$, are control points, where $P_i\in R^3$. Then the n-rank Bézier curve generated by these $n+1$ points is defined as follows:

$$\mathrm{P}(u)=\sum_{i=0}^{n}P_i\cdot B_{i,n}(u)\ ,\ \ u\in[0,1] \tag{1}$$

here $B_{i,n}(u)=C_n^i\cdot(1-u)^{n-i}\cdot u^i$ is called Bernstein basic function. When the parameter u changes from 0 to 1, we can get a curve generated by formula (1) which is inside the polygon constituted by control points. The curve is called Bézier curve generated by the control polygon.

In the above processing, the control points are all geometric points in three-dimension space, and the curve is also in the polygon generated by control points. Now let us consider an extended situation. If the control points are nonobjective numeral points, then the points on the curve generated by formula (1) will also be nonobjective numeral points, and we can get a nonobjective numeral curve generated by formula (1) which is inside the nonobjective numeral polygon constituted by the nonobjective numeral control points. When the control points are random then we can get a random curve generated by the random control points.

2.2 Extended Chaotic Sequence Generating Method

Based on the statement in section 2.1, we present the extended chaotic sequence generating method based on the above analysis.

Suppose $\{a_1^i, a_2^i, \cdots, a_n^i, \cdots\}$, （$i = 0, 1, \cdots, \mathrm{K}$） are known chaotic sequences, and suppose the sequences have the same dimension D and $\{t_1, t_2, \cdots\}$ is a chaotic sequence with one or D dimension. Then we can get an extended chaotic sequence $\{a_1, a_2, \cdots\}$ as follows:

$$a_n = \sum_{i=0}^{K} a_n^i \cdot B_{i,n}(t_n) \ , \ n = 1, 2, \cdots \tag{2}$$

when $\{t_1, t_2, \cdots\}$ is one dimension or

$$a_n^j = \sum_{i=0}^{K} a_n^{i,j} \cdot B_{i,n}(t_n^j) \ , \ n = 1, 2, \cdots \ \ j = 1,2, \cdots, D \tag{3}$$

when $\{t_1, t_2, \cdots\}$ is D dimensions.

For example, we can show an extended chaotic sequence generated by Arnold chaotic system for $D = 2$. Arnold chaotic system is defined as:

$$\begin{cases} x_{n+1} = (x_n + y_n) \mod 1 \\ y_{n+1} = (x_n + 2y_n) \mod 1 \end{cases} \tag{4}$$

where mod 1 is defined as: $x \bmod 1 = x - \lfloor x \rfloor$, so the sequence generated by formula (4) is inside the square $[0,1] \times [0,1]$. We can describe the formula (4) in matrix form:

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} x_n \\ y_n \end{bmatrix} \bmod 1 \tag{5}$$

If we generate the control points and the sequence $\{t_1, t_2, \cdots\}$ using formula (4) and Arnold chaotic system, then with formula (3), we can get an extended chaotic sequence.

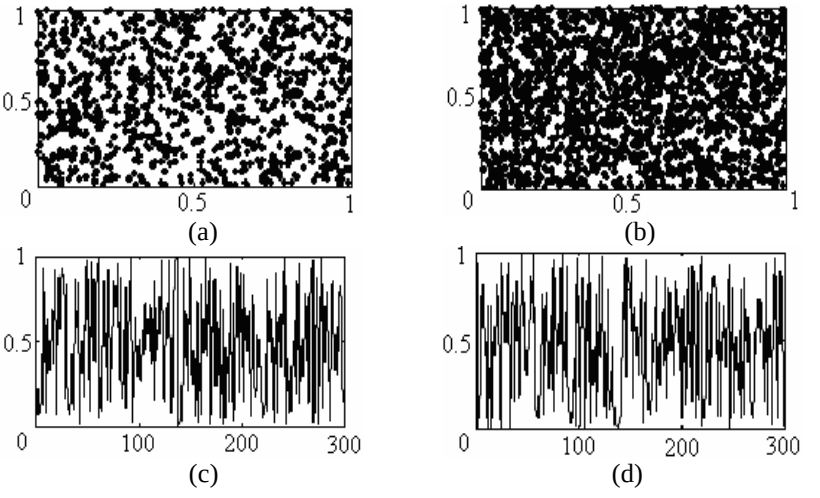


Fig. 1. Sequences generated by Arnold system: (a) generating 1000 points (b) generating 2000 points (c) x coordinator's randomness analysis (d) y coordinator's randomness analysis

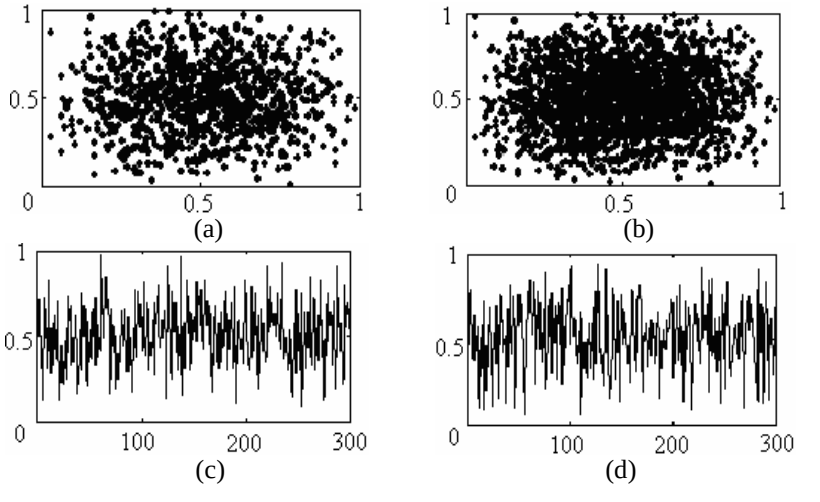


Fig. 2. Sequences generated by formula(3): (a) generating 1000 points (b) generating 2000 points (c) x coordinator's randomness analysis (d) y coordinator's randomness analysis

Fig.1 and Fig.2 give the simulation results for Arnold chaotic system and our method, it is shown that the sequences generated by formula(4) are random and proportional distributing in $[0,1]\times[0,1]$, and the sequences generated by formula (3) are random and its distribution is approximately inside a disc whose center is $(0.5, 0.5)$. It should be pointed out that for image's two-dimension gray histogram[3,4], the most pixels inside the target class and the background classes' distribution are distributed

near the diagonal from $(1,1)$ to (L,L), this property means that the most thresholds for image segmentation are distributed near the diagonal from $(1,1)$ to (L,L). So the sequence generated by formula(3) is superior to the Arnold chaotic sequence for image segmenting in the case of two-dimension gray histogram.

3 Two-Dimension Otsu's Image Segmentation Method Based on Chaotic Optimization

Otsu's method[2] is an attractive threshold selection method due to its simplicity and effective. In Otsu's paper, the between-class variance (BCV) is defined, and the gray level at the BCV maximum determines an optimal threshold. But in practice, one-dimension Otsu's method would fail to apply in cases of images with noises. Liu and Li[3] extended Otsu's method to two-dimension case, the detail is stated in the following.

Let the image's gray levels is L, then the pixels' neighborhood's gray levels are also L. For every pixel, we can compute its average gray value, with these values, for every point (i,j) where i is the pixel's gray value and j is the pixel's neighborhood's average gray value, the frequencies of (i,j) is written as f_{ij}, then we can define the joint probability of (i,j) as: $P_{ij}=\dfrac{f_{ij}}{M\times N}$, where $M\times N$ is the size of the image. We have $M\times N=\sum_{i=1}^{L}\sum_{j=1}^{L}f_{ij}$, $\sum_{i=1}^{L}\sum_{j=1}^{L}P_{ij}=1$.

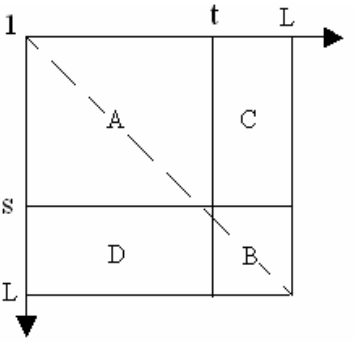


Fig. 3. Two-dimension histogram

For two-dimension gray histogram, there are L^2 elements. On the assumption that (s,t) is the segmentation threshold, the two-dimension histogram is divided into four parts just as Fig.3 shows. Because the most pixels inside the target class and the background classes are consistent, so they are distributed near the diagonal from $(1,1)$ to (L,L). That means region A and B involve most pixels of target class and background classe, region C and D involve most pixels of edge and the noises.

Now let:

$$P_A(s,t) = P(A) = \sum_{i=1}^{s}\sum_{j=1}^{t} P_{ij}\ , P_B(s,t) = P(B) = \sum_{i=s+1}^{L}\sum_{j=t+1}^{L} P_{ij} \tag{6}$$

The mean values of region A and B are computed as follows:

$$\begin{aligned}\bar{u}_A = (u_{Ai}, u_{Aj})^T &= \left(\sum_{i=1}^{s} P(i|A), \sum_{j=1}^{t} P(j|A)\right)^T \\ &= \left(\sum_{i=1}^{s}\sum_{j=1}^{t} \frac{i\cdot P_{ij}}{P_A(s,t)}, \sum_{i=1}^{s}\sum_{j=1}^{t} \frac{j\cdot P_{ij}}{P_A(s,t)}\right)^T \\ \bar{u}_B = (u_{Bi}, u_{Bj})^T &= \left(\sum_{i=s+1}^{L} P(i|B), \sum_{j=t+1}^{L} P(j|B)\right)^T \\ &= \left(\sum_{i=s+1}^{L}\sum_{j=t+1}^{L} \frac{i\cdot P_{ij}}{P_B(s,t)}, \sum_{i=1}^{s}\sum_{j=1}^{t} \frac{j\cdot P_{ij}}{P_B(s,t)}\right)^T\end{aligned} \tag{7}$$

the whole mean value is computed as:

$$\bar{u} = (u_i, u_j)^T = \left(\sum_{i=1}^{L}\sum_{j=1}^{L} i\cdot P_{ij}, \sum_{i=1}^{L}\sum_{j=1}^{L} j\cdot P_{ij}\right)^T \tag{8}$$

Now defining the scatter matrix as follow:

$$S_B(s,t) = P_A(s,t)\cdot[(\bar{u}_A - \bar{u})\cdot(\bar{u}_A - \bar{u})^T] + P_B(s,t)\cdot[(\bar{u}_B - \bar{u})\cdot(\bar{u}_B - \bar{u})^T] \tag{9}$$

The trace of $S_B(s,t)$ is defined as:

$$Tr(S_B(s,t)) =$$

$$P_A(s,t)\cdot[(u_{Ai} - u_i)^2 + (u_{Aj} - u_j)^2] + P_B(s,t)\cdot[(u_{Bi} - u_i)^2 + (u_{Bj} - u_j)^2] \tag{10}$$

The optimal threshold value (s^*, t^*) is obtained by:

$$(s^*, t^*) = \underset{1\le s\le L,\, 1\le t\le L}{Arg}\ \{Tr(S_B(s,t))\} \tag{11}$$

In many cases, pixels in region C and D can be ignored. So we can assume that

$$P_A(s,t) + P_B(s,t) \approx 1$$

then we have

$$\bar{u} \approx P_A(s,t)\cdot\bar{u}_A + P_B(s,t)\cdot\bar{u}_B$$

Under this assume, a more simple expression is given by Liu and Li[3].

$$Tr(S_B(s,t)) = \frac{[(P_A(s,t)\cdot u_i - u_{Zi})^2 + (P_A(s,t)\cdot u_j - u_{Zj})^2]}{P_A(s,t)\cdot P_B(s,t)} \tag{12}$$

where $u_{Zi} = \sum_{i=1}^{s}\sum_{j=1}^{t} i\cdot p_{ij}$, $u_{Zj} = \sum_{i=1}^{s}\sum_{j=1}^{t} j\cdot p_{ij}$.

Liu and Li's method is validity for noise image, but the computation time is very large, to reduce the computation amount, Gong et.al[4] gave a recursion algorithm using formula (12). In practice, the supposition that "the sum of the target and the background's probability is approximate to 1" is not fit in some cases, thus we consider to use formula (10) not to use formula (12) as the selection criterion. But now we are confronted with the tremendous computing complexity. How to solve this problem? In the following, we will use chaotic optimization method to determine the good threshold value. The algorithm is described as follows:

Step1: Input the image , the times of iterative p , the number of the initial points r and the initial conditions of chaotic system. Set the gray levels region: $\{1,2,\cdots,L\}\times\{1,2,\cdots,L\}$.

Step2: Generating the chaotic sequences $(x_1,y_1),(x_2,y_2),\cdots,(x_r,y_r)$ with formula (4) and (3), discrete the sequences from real number region into integer region. Get the sequences $(X_1,Y_1),(X_2,Y_2),\cdots,(X_r,Y_r)$.

Step3: Computing the $Tr(S_B(X_i,Y_i))$ by formula(10), then generate the new chaotic sequences with formula (4) and (3), discrete the sequences from real number region into integer region, at the same time, computing the $Tr(S_B(X_i,Y_i))$ of new chaotic sequences, comparing the new $Tr(S_B(X_i,Y_i))$ with the previous iterative results. Reserve the front r maximum values $Tr(S_B(X_i,Y_i))$ from the obtained $2r$ $Tr(S_B(X_i,Y_i))$ values and its corresponding gray values (X_i,Y_i) .

Step4: Continue the computation process for p times, then find the satisfied result in the last r points (X_i,Y_i) .

In step 4 of the above algorithm, the words "satisfied result" relies on the man and the case the image may be used, it may be subjective or objective. For example, we can get a better segmenting result from the last r results based on vision effect. In this paper, in order to compare the efficiency of our method with existed two-dimension Otsu's method [3,4], we think that the "satisfied result" is the (X_i,Y_i) corresponding to the maximum value of $Tr(S_B(X_i,Y_i))$, this can be consider as a approximate solution for formula(11).

Now we give the simulation results based on Arnold system and the 3×3 neighborhood. As compare, we also present the results of the two-dimension Otsu's method using the formula (10) (without using the hypothesis that "the sum of the target and the background's probability is approximate to 1") by exhaust search, the two-dimension Otsu's recursion algorithm using the formula (12) and the chaotic

optimization of Otsu's image segmentation method in two-dimension gray histogram using the formula (10) and the Arnold chaotic sequence.

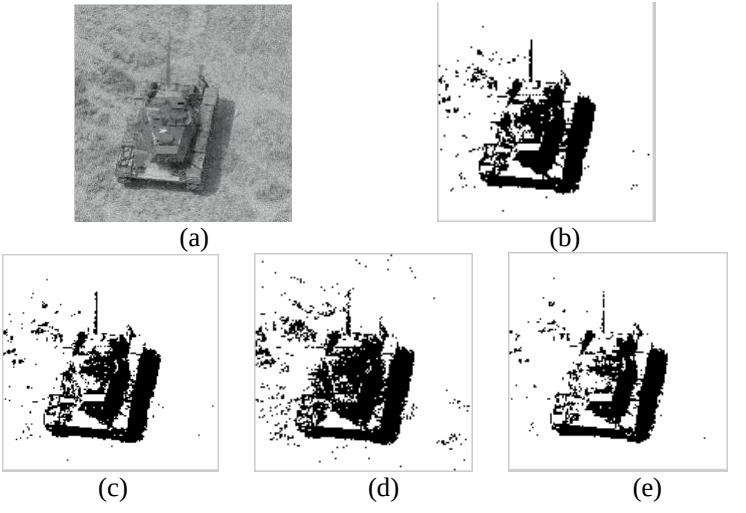


Fig. 4. Results of image segmentation (a) Image with noises (b) Results of chaotic optimization using the expended chaotic sequence (c) Results of two-dimension Otsu's algorithm using formula (10) (d) Results of recursion algorithm using formula (12) (e) Results of chaotic optimization using the Arnold sequence

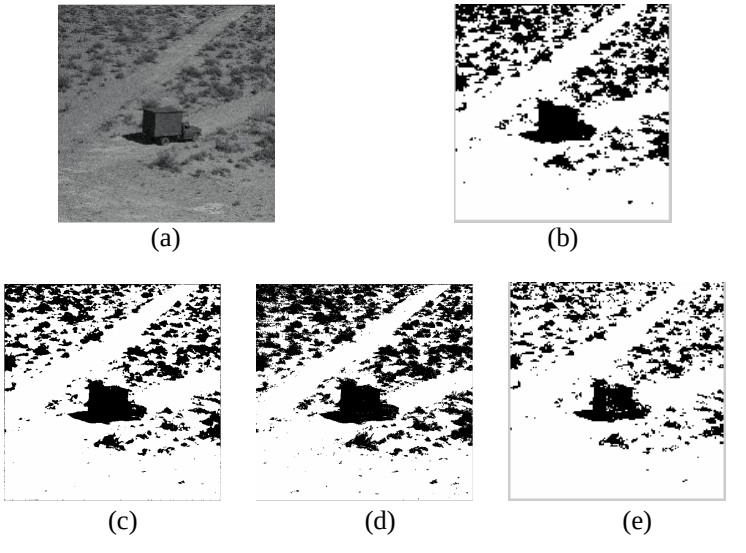


Fig. 5. Results of image segmentation (a) Image with noises (b) Results of chaotic optimization using the expended chaotic sequence (c) Results of two-dimension Otsu's algorithm using formula (10) (d) Results of recursion algorithm using formula (12) (e) Results of chaotic optimization using the Arnold sequence

Table 1 shows the performances of the four methods, where T means runtime the method expended and the unit of T is second. H means the threshold the method got. Tr means the values of $Tr(S_B(s,t))$ corresponding to the threshold we have got.

Table 1. Performances of four methods

		Fig 4	Fig 5
Chaotic optimization using extended chaotic sequence	T	1.6520	1.6120
	H	(144,109)	(88,119)
	Tr	1.1722e+003	1.1488e+003
Two-dimension Otsu's method	T	146.0200	145.8500
	H	(141,106)	(118, 88)
	Tr	1.1735e+003	1.2044e+003
Recursion method	T	1.5220	1.5520
	H	(118,121)	(100,102)
	Tr	1.0147e+003	997.4332
Chaotic optimization using Arnold chaotic sequence	T	1.6830	1.6920
	H	(141, 103)	(80,120)
	Tr	1.1656e+003	1.1293e+003

From the simulation results and table 1, we can seen that the segmentation results using our method is approximately equivalent to the two-dimension Otsu's method using formula (10), but the efficiency of our method is better than the two-dimension Otsu's method using formula (10) because our method optimize the computing process.

On the other hand, the efficiency of our method is comparative to the recursion method using formula (12), but the segmentation results using our method is better than the results of the recursion method using formula (12) because the computation process of our method is accurate without using the hypothesis that "the sum of the target and the background's probability is approximate to 1".

If we set the value of $Tr(S_B(X_i,Y_i))$ as the optimization target value, from the simulation results we know that the effect of chaotic optimization method using the extended chaotic sequences is better than the method using the Arnold chaotic sequences. This means that our method is more fit to image segmentation.

4 Conclusion

In this paper, we present a chaotic sequence generating method. The simulation results show that the sequence generated by our method are pseudorandom, and the most important characteristic of the extended chaotic sequence is that its distribution is approximately inside a disc whose center is $(0.5, 0.5)$, this characteristic indicates that the sequence is fit to image segmentation in two-dimension gray histogram. As an application, we also present a two-dimension Otsu's image segmentation method based on chaotic optimization. If we set the value of $Tr(S_B(X_i,Y_i))$ as the optimization target value, The simulation results show that the effect using our

method is approximately equivalent to the two-dimension Otsu's method using formula (10) and the efficiency using our method is comparative to the recursion method using formula (12). In the experiment we used a objective criterion. In the future work, we will consider the case using subjective criterion.

Acknowledgement

The work is supported by National Nature Science Funds project (Grant Number: 60572133) and Opening Project of Computer Network and Information Security Laboratory(CNIS) of Ministry of Education.

Reference

[1] Lee SU, Chung SY (1990) A comparative performance study of several global thresholding techniques for segmentation[J]. Comput Graph Image Process 52:171 — 190.
[2] Otsu N(1979). A threshold selection method from gray-level histograms[J]. IEEE Trans on Systems, Man and Cybernetics. 9 (1): 62-66.
[3] Liu Jian-zhuang, Li Wen-qing(1993). Two-dimension Otsu's Automatic Segmentation method of Gray Image[J]. Automatization Journal. 19 (1): 101-105(in Chinese)
[4] Gong J, Li L Y, Chen W N(1998). Fast recursive algorithm for two-dimensional thresholding[J]. Pattern Recognition. 31(3): 295-300.
[5] Han Hen and Neri Merhav(2004). On the Threshold Effect in the Estimation of Chaotic Sequences[J]. IEEE Trans on Information Theory. 50(11): 2894-2904
[6] L. Wang and K. Smith(1998). On chaotic simulated annealing[J]. IEEE Trans on Neural Networks. 9(4): 716-718
[7] Claudio R. Mirasso et.al (2002). Chaos Shift-Keying Encryption in Chaotic External-Cavity Semiconductor Lasers Using a Single-Receiver Scheme[J]. IEEE Photonics Technology Letters. 14(4): 456-458
[8] Liang Zhao et.al (2002). A Network of Globally Coupled Chaotic Maps for Adaptive Multi-Resolution Image Segmentation[C]. Proceedings of the VII Brazilian Symposium on Neural Networks(SBRN'02)
[9] Xiu Chun-bo et.al(2004). Optimal Entropy Thresholding Image Segmentation Based on Chaos Optimization[J]. Computer Engineering and Applications Journal. 2004(27): 76-77(in Chinese)
[10] Les Piegl, Wayne Tiller(1995). The NURBS Book[M]. Springer.

A Study on Construction of Time-Varying Orthogonal Wavelets*

Guangming Shi[1], Yafang Sun[1], Danhua Liu[1], and Jin Pan[2]

[1] School of Electronic Engineering, Xidian University, Xi'an, 710071, China
gmshi@xidian.edn.cn, yfsun@mail.xidian.edu.cn
[2] Lab of Network Security and Countermeasure, Xi'an Communications Institute, Xi'an, 710106, China

Abstract. Time-varying wavelets are highly desired in exploiting the nonstationarity of signals. However, it is difficult to hold the perfect reconstruction (PR) and regularity properties simultaneously in the construction of time-varying wavelets. This paper proposes a simple method to construct time-varying orthogonal wavelets based on the lattice structure of two-channel paraunitary (PU) filter banks, in which both the PR and orthogonality properties are well preserved. The regularity conditions imposed on the lattice structure are expressed in terms of the lattice coefficients and the wavelet filter banks are obtained by using an optimization technique. Then the time-varying orthogonal wavelets can be constructed by the lattice structure formulation for time-varying filter banks. Design examples show that this method is of great flexibility and effectiveness.

1 Introduction

In the field of signal and image processing, time-varying filter banks and time-varying wavelet transforms have attracted considerable attention as powerful tools to exploit the non-stationary of signals[1]-[4]. So far there are various construction techniques for time varying filter banks [2], [5]-[17]. Furthermore, many different approaches have been proposed to resolve the perfect reconstruction (PR) problem of switching between two sets of filters on both the analysis and synthesis sides of the filter bank structure. Two prominent types are mentioned here. One is designing transition filters [5], [7], [11]-[14] in the transition period to maintain the PR property, and the other is obtaining PR time-varying filter banks by cascading sections of FIR lattice [18] with time-varying parameters [6], [8], [16]. It is well known that the wavelet bases can be obtained from PR filter banks with added regularity conditions [19]-[26]. Similarly, time-varying wavelets can be constructed by using time-varying filter banks. In [12], the authors proposed two methods to implement time-varying wavelet transforms. One employs a two-band PR time-varying filter bank in the tree structure and the other is based on the time-varying parallel structure. The former achieves good reconstruction performance but requires intermediate transforms when changing from one wavelet transform to another. On the contrary, the latter provides immediate

* Supported by National Natural Science Foundation of China (Grant No. 60372047).

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 284 – 294, 2006.

switching. However, the reconstruction is much more difficult. In [14], time-varying discrete wavelet bases are obtained by taking of the effective input/output map of an arbitrary unitary time-varying filter bank tree. Time-varying wavelet transforms with lifting steps were introduced in [2]. The corresponding time-varying filter bank is switched between three filter banks, depending on the standard deviation estimated by the maximum likelihood problem. In [27], two-channel wavelet filter banks with adaptive number of zero moments based on the lifting scheme were proposed. The number of zero moments is changed at each step of decomposition or reconstruction.

Time variant wavelet filter banks are more suitable for analysis of non-stationary signals than fixed banks. But up to now, the problem that maintaining both the PR and regularity properties simultaneously has not been effectively resolved. In this paper, a new method is proposed to construct time-varying orthogonal wavelets based on the lattice structure, which possesses the PR property structurally. The regularity conditions of two-channel lattice-structured paraunitary (PU) filter banks are discussed, and the relationship among the lattice coefficients is also pointed out. Consequently, an effective optimization algorithm with constraints is employed to design wavelet filter banks. Moreover, we extend this structure to time-varying case to obtain time-varying orthogonal wavelets. Design examples show the flexibility and the effectiveness of the presented method.

This paper is organized as follows. Section 2 reviews the lattice structure of two-channel PU filter banks. Section 3 gives the relation between lattice coefficients and the regularity, and discusses how to design time-varying wavelets based on the lattice structure. Section 4 presents the design procedure and several examples. Finally some conclusions are drawn in Section 5.

2 Two-Channel Lattice-Structured PU Filter Banks

The lattice structure is an attractive structure in the design of PU filter banks [18]. In Fig. 1, a lattice structure of J stages is given, where $J = N/2$ and N is even, denoting the length of filters. $\{P_{2m-1}(z), Q_{2m-1}(z)\}$, $m = 1, \cdots, N/2$ is defined as the

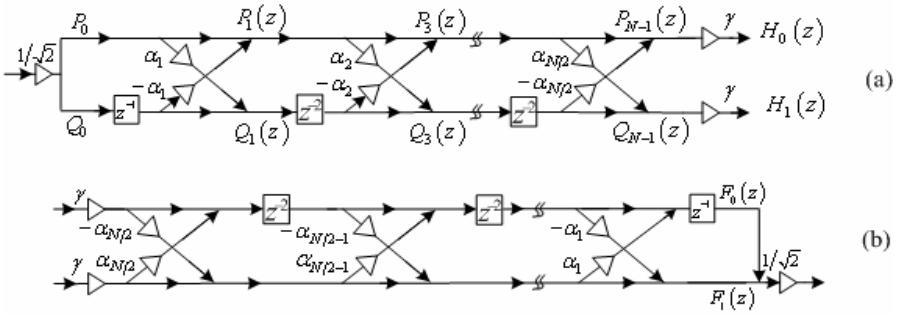


Fig. 1. The lattice structure. (a) The analysis bank, and (b) the synthesis bank.

transfer function pair of the *m*-th lattice section. Then the relation between the neighbor lattice sections is given by

$$\begin{aligned} P_{2m-1}(z) &= P_{2m-3}(z) - \alpha_m z^{-2} Q_{2m-3}(z) \\ Q_{2m-1}(z) &= \alpha_m P_{2m-3}(z) + z^{-2} Q_{2m-3}(z), \end{aligned} \tag{1}$$

where α_i denote lattice coefficients. So that the corresponding lowpass and highpass filters of the analysis bank are given by $H_0(z) = \gamma P_{N-1}(z)$ and $H_1(z) = \gamma Q_{N-1}(z)$, respectively, where $\gamma^2 = \prod_{i=1}^{J} \left(1+\alpha_i^2\right)^{-1}$. Consequently, $H_0(z)$ and $H_1(z)$ satisfy the "power complementary property"

$$\left|H_0\left(e^{j\omega}\right)\right|^2 + \left|H_1\left(e^{j\omega}\right)\right|^2 = d\,, \tag{2}$$

where d is some constant. In Fig. 1, $d=1$, and the relation between $H_0(z)$ and $H_1(z)$ is expressed as

$$H_1(z) = z^{-(N-1)} H_0\left(-z^{-1}\right). \tag{3}$$

And the two corresponding synthesis filters $F_0(z)$ and $F_1(z)$ are given by $F_0(z) = z^{-(N-1)} H_0\left(z^{-1}\right)$ and $F_1(z) = z^{-(N-1)} H_1\left(z^{-1}\right)$.

As discussed in [28], $\left\{H_0(z), H_1(z)\right\}$ satisfies the PR condition if all $\left\{P_{2m-1}(z), Q_{2m-1}(z)\right\}$, $m = 1, \cdots, N/2$ satisfy the power-complementary image pair condition. Hence, if we remove the rightmost lattice section shown in Fig. 1(a), we do not destroy the PR property but only form a new system with a deterioration of the analysis filters performance. Similar comment also holds for the synthesis bank.

The polyphase matrix $\mathbf{E}(\mathbf{z})$ of a PU filter bank of degree J can be factorized as

$$\mathbf{E}(\mathbf{z}) = \lambda \mathbf{R}_J \Lambda(z) \mathbf{R}_{J-1} \cdots \Lambda(z) \mathbf{R}_1\,, \tag{4}$$

where $\lambda = \gamma/\sqrt{2}$, $\mathbf{R}_m = \begin{pmatrix} 1 & -\alpha_m \\ \alpha_m & 1 \end{pmatrix}$, $m = 1, 2, \cdots, N/2$ and $\Lambda(z) = \begin{pmatrix} 1 & 0 \\ 0 & z^{-1} \end{pmatrix}$. Therefore, $\mathbf{E}(\mathbf{z})$ can obtain orthogonality if each lattice section is orthogonal. That is to say, we can construct orthogonal wavelets using the lattice structure. This will be discussed in detail in the latter section.

3 Time-Varying Orthogonal Wavelets Based on Lattice Structures

In this section, we focus on two problems: the *K*-regular conditions corresponding to the lattice coefficients and the construction of time-varying wavelets in the lattice structure.

It has known that two-channel FIR PU filter banks satisfying the *K*-regular conditions can be used to generate orthogonal wavelets. Hence, time-varying

orthogonal wavelets can be obtained by employing the lattice structure based on two-channel time-varying FIR filter banks. In [29], the theorem which guarantees the regularity has been proved; it is also applicable to the two-channel case. In two-channel filter banks, it is simplified as follows:

Corollary: For two-channel PU filter banks, $H_1(z)$ has K zeros at $z=1$, if and only if $H_0(z)$ has K zeros at $z=-1$. Equivalently,

$$\frac{d^n}{dz^n}\mathbf{E}(\mathbf{z}^2)\begin{bmatrix}1\\ z^{-1}\end{bmatrix}\Bigg|_{z=1}=\begin{bmatrix}c\\ 0\end{bmatrix},\qquad n=0,1,\cdots,K-1, \tag{5}$$

where c is a non-zero element.

3.1 The *K*-Regular Conditions of Lattice-Structured Filter Banks

Here, according to the polyphase matrix $\mathbf{E}(\mathbf{z})$ expressed in (2), we will study the regularity conditions of the lattice structure for $K=1$, 2, and 3 as examples.

In the case of $K=1$: Here the condition on the lattice structure is derived such that the resulting PU filter banks have one vanishing moment. Substituting $n=0$ into (5), we get

$$\lambda\mathbf{R}_J\mathbf{R}_{J-1}\cdots\mathbf{R}_1\begin{bmatrix}1\\1\end{bmatrix}=\begin{bmatrix}1\\0\end{bmatrix}. \tag{6}$$

This equation (6) provides a necessary and sufficient condition for the filter banks having one degree of regularity.

In the case of $K=2$: Similarly, we can deduce the necessary and sufficient condition for the filter banks to possess two vanishing moments. Substituting $n=1$ into (5), it can be simplified to

$$\begin{aligned}-2\lambda\big[&\mathbf{R}_J\Psi\mathbf{R}_{J-1}\mathbf{R}_{J-2}\cdots\mathbf{R}_1+\mathbf{R}_J\mathbf{R}_{J-1}\Psi\mathbf{R}_{J-2}\cdots\mathbf{R}_1+\cdots\\ &+\mathbf{R}_J\mathbf{R}_{J-1}\cdots\mathbf{R}_2\Psi\mathbf{R}_1\big]\begin{bmatrix}1\\1\end{bmatrix}-\lambda\mathbf{R}_J\cdots\mathbf{R}_1\begin{bmatrix}0\\1\end{bmatrix}=\begin{bmatrix}c\\0\end{bmatrix},\end{aligned} \tag{7}$$

where $\Psi=\begin{pmatrix}0&0\\0&1\end{pmatrix}$ and c is an uncertain constant. It should be noticed that in order to have two vanishing moments for the filter bank, equation (5) must be valid for both $n=0$ and $n=1$. That is to say, conditions (6) and (7) must be satisfied simultaneously and they are also the necessary and sufficient conditions.

In the case of $K=3$: For the further study, here we also give the condition for three vanishing moments. Substituting $n=2$ into (5) suggests that

$$
\begin{aligned}
2\lambda\{&\left[\left(3\mathbf{R}_J\Psi\mathbf{R}_{J-1}\cdots\mathbf{R}_1+2\mathbf{R}_J\Psi\mathbf{R}_{J-1}\Psi\mathbf{R}_{J-2}\cdots\mathbf{R}_1+\cdots+2\mathbf{R}_J\Psi\mathbf{R}_{J-1}\cdots\mathbf{R}_2\Psi\mathbf{R}_1\right)\right.\\
&+\left(2\mathbf{R}_J\Psi\mathbf{R}_{J-1}\Psi\mathbf{R}_{J-2}\cdots\mathbf{R}_1+3\mathbf{R}_J\mathbf{R}_{J-1}\Psi\mathbf{R}_{J-2}\cdots\mathbf{R}_1+\cdots+2\mathbf{R}_J\mathbf{R}_{J-1}\Psi\mathbf{R}_{J-2}\cdots\mathbf{R}_2\Psi\mathbf{R}_1\right)\\
&\left.+\cdots+\left(2\mathbf{R}_J\Psi\mathbf{R}_{J-1}\cdots\mathbf{R}_2\Psi\mathbf{R}_1+\cdots+3\mathbf{R}_J\cdots\mathbf{R}_2\Psi\mathbf{R}_1\right)\right]\mathbf{e}+2\left(\mathbf{R}_J\Psi\mathbf{R}_{J-1}\cdots\mathbf{R}_1\right.\\
&\left.+\mathbf{R}_J\mathbf{R}_{J-1}\Psi\mathbf{R}_{J-2}\cdots\mathbf{R}_1+\cdots+\mathbf{R}_J\mathbf{R}_{J-1}\cdots\mathbf{R}_2\Psi\mathbf{R}_1\right)\mathbf{v}+\mathbf{R}_J\mathbf{R}_{J-1}\cdots\mathbf{R}_1\mathbf{v}\}=\begin{bmatrix}c\\0\end{bmatrix},
\end{aligned}
\tag{8}
$$

for some constant c, where $\mathbf{e}=\begin{bmatrix}1\\1\end{bmatrix}$, $\mathbf{v}=\begin{bmatrix}0\\1\end{bmatrix}$, Ψ is defined as above. Note that if the filter banks have three vanishing moments, it is necessary to satisfy all conditions (6), (7) and (8). Since the conditions for higher degree of regularities are much more complex, in this paper, we only study the case with three regularities.

At the time being, we have obtained the conditions for three regularities. And then using the optimization algorithm with the regularity conditions as constraints, we can design PU filter banks having one or two or three vanishing moments. Moreover, the desired orthogonal wavelets can be obtained by the iteration of filter banks designed above. Two examples are displayed in section 4.

3.2 The Time-Varying Wavelet Based on Lattice Structure

In view of the relation between wavelets and filter banks, it is suggested that changing the wavelets is equivalent to changing the analysis section in the lattice structure. In time-varying systems, how to preserve the PR property during the transition period is an important problem. In this section, an efficient method for switching between two sets of lattice coefficients is introduced. From the switching process, it will show that there are transition functions when changing from one wavelet to another.

The lattice structure applied to time-varying banks had been introduced in [6]. To obtain a PR time-varying system, we can change one or all of the lattice sections in the analysis bank, and ensure the corresponding synthesis section to change at the correct time. For example, there are J lattice sections in the analysis bank in Fig. 1(a). If we change the i-th lattice section at time n_0, the corresponding section in the synthesis bank must be changed at the time $n_0+(J-i)$. Here we also simply show how to switch two different sets of lattice coefficients in the transition period. Given two sets lattice coefficients $\alpha_1,\cdots,\alpha_J$ and $\beta_1,\cdots,\beta_J$, assume the switch takes place at time n_0, then replacing α_1 with β_1 at this time while " turn off " α_J through α_2 and replacing each of them with a zero. The remaining β's will be "turned on" in a cascade fashion at different time, until all α's are replaced with β's, the time-varying procedure is performed. For many cases, the switching method makes the transitions somewhat smooth.

According to the hierarchical property of the lattice structure, if we delete the rightmost lattice section, a new system with analysis and synthesis filters of reduced length $N-2$ will be generated, and the polyphase matrixes for the filters are still paraunitary. Hence, we can obtain time-varying wavelets through changing the regularity and the length of the wavelet filter banks.

4 Design Procedure and Examples

In this section, we discuss how to design wavelet filter banks and construct time-varying wavelets using the lattice structure. In addition, some examples are also presented. The *K*-regular conditions referred to lattice coefficients have been given in section 3, now we can use optimization algorithm with constraints to design the wavelet filter banks. Since the two-channel filter bank is orthogonal, we only need to design the lowpass filter $H_0(z)$. The objective function is given by

$$\Phi = \int_{\omega_s}^{\pi} \left| H_0\left(e^{j\omega}\right) \right|^2 d\omega, \tag{9}$$

where the parameter ω_s denotes the stopband frequency of $H_0\left(e^{j\omega}\right)$. The objective function is used to minimize the stopband attenuation of $H_0(z)$, which is sufficient to guarantee good frequency response. In this paper, the regularity conditions are used as constraints. The design procedure is described as follows.

Step 1 Determine the length N of the filters, and N is even.

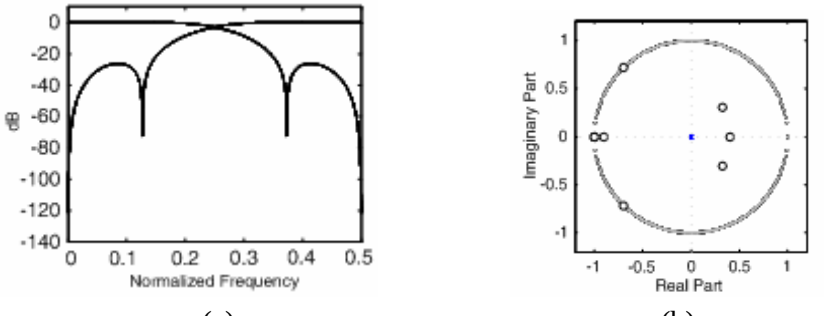
Step 2 Obtain $N/2$ initial estimates of the coefficients α_m, $m = 1, \cdots, N/2$ using the method introduced in [28].

Step 3 Calculate the objective function in terms of the relation between α_m and $H_0(z)$.

Step 4 Select the condition of the *K*-regularity as constraint, here $K = 1, 2, 3$.

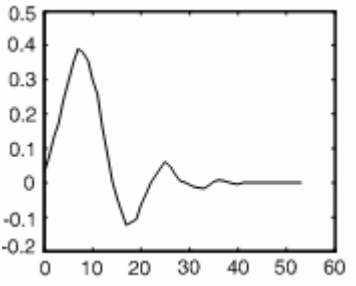
Step 5 Take the initial estimates of α_m as inputs to the quasi-Newton routine and optimize the objective function using the function *fmincon.*

Design Example 1: In this example, a 1-regular 8-tap lattice-structured filter bank is designed using the optimization algorithm. Fig. 2(a) shows the frequency responses of the resulting analysis filters. The zeros of the lowpass filter plotted in Fig. 2(b) show

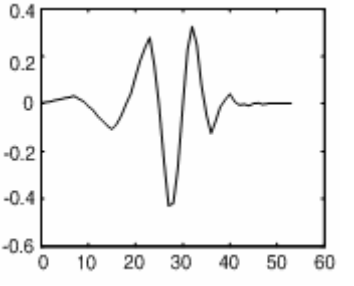


(a) (b)

Fig. 2. A 1-regular PU filter bank with length 8. (a) frequency responses, (b) zero locations of the lowpass filter, (c) the scaling function, and (d) the wavelet function.



(c)

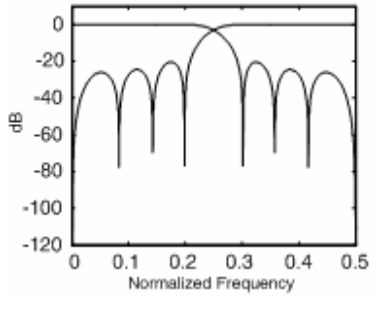


(d)

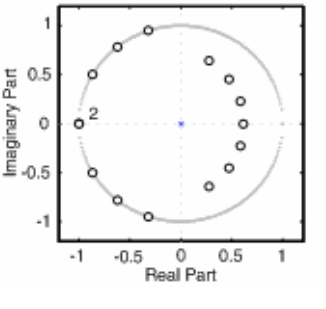
Fig. 2. (*Continued*)

that the filter bank has one degree of regularity which confirms the theory. The corresponding scaling and wavelet function are shown in Fig. 2(c) and (d), respectively.

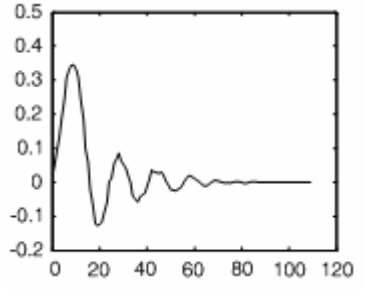
Design Example 2: In this example, a 2-regular 16-tap lattice-structured filter bank is designed. Fig. 3(a) shows the frequency responses of the resulting analysis filters, the



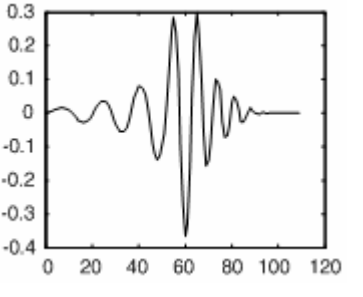
(a)



(b)



(c)



(d)

Fig. 3. A 2-regular PU filter bank with length 16. (a) frequency responses, (b) zero locations of the lowpass filter, (c) the scaling function, and (d) the wavelet function.

zeros of the lowpass filter are plotted in Fig. 3(b). It is known that the highpass filter will have two vanishing moments because the lowpass filter has two zeros at the mirror frequency. The corresponding scaling and wavelet function are shown in Fig. 3(c) and (d), respectively.

Design Example 3: A switching process between two wavelets is given in this example. The wavelet designed in example 1 is called the old wavelet, and the wavelet in example 2 is the new one. They are shown in Fig. 4(a) and (c), respectively. The

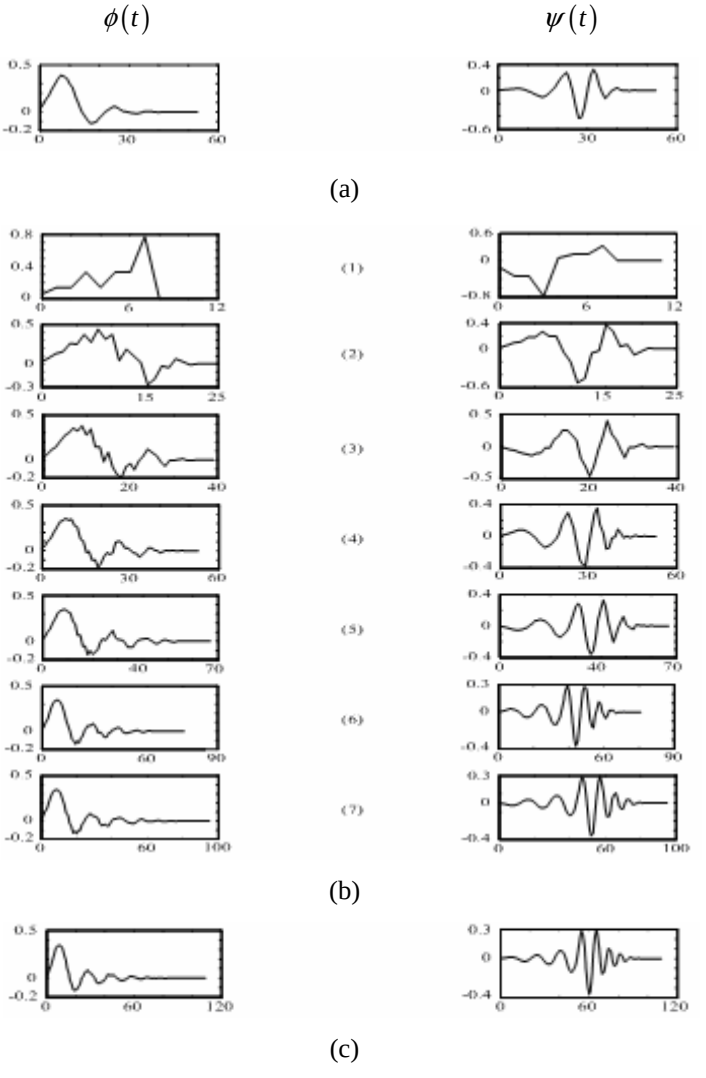


Fig. 4. A complete process of switching from the wavelet with one vanishing moment (*shown in Fig. 2(d)*) to the one with two vanishing moments (*shown in Fig. 3(d)*). (a) The scaling and wavelet functions of the old wavelet, (b) the seven switching states, and (c) the scaling and wavelet functions of the new wavelet.

switching between these two wavelets is achieved by using the time-varying lattice structure, and the whole process is shown in Fig. 4(b). It can be seen that there are seven transition states before the new wavelet is formed. The scaling functions and the corresponding transitions are shown in the left column and the wavelets and the corresponding transitions are shown in the right column.

The simulations in Fig. 4 show that changing from one wavelet to another using the lattice structure requires intermediate functions, but transitions are relatively smooth. And in view of the switching method, the transitions are only associated with the new scaling and wavelet functions.

We also have tried to obtain the initial estimates of α_m by starting from a half-band filter [18], but it is subject to severe numerical errors to find a spectral factor of such a half-band filter. In addition, the nonlinear optimization technique which we selected needs relatively small number of iterations for convergence, though each of these iterations might be more time consuming than that of other methods such as conjugate-gradient technique. Therefore, the desired wavelets can be obtained easily by using the proposed method.

5 Conclusion

In this paper, the construction of time-varying wavelets based on the lattice structure of two-channel PU filter banks has been studied. The proposed method guarantees that the wavelets have not only the one or two or three vanishing moments but also the orthogonal property. Furthermore, we can switch the wavelets generated by filter banks with different lengths and regularities. And the switching method makes the transitions somewhat smooth. Simulations show that our approach is flexible and efficient to construct time-varying orthogonal wavelets.

References

1. Xiong, Z.X., Ramchandran, K., Herley, C., Orchard, M.T.: Flexible Tree-Structured Signal Expansions Using Time-Varying Wavelet Packets. IEEE Trans. Signal Processing, Vol. 45, No. 2, (1997) 333–345
2. Okuda, M., Mitra, S.K., Ikehara, M., Takahashi, S.: Time-Varying Wavelet Transforms with Lifting Steps for Lossless Image Compression. Proc. IEEE Int. Conf. Image Processing, Vol. 4, (1999) 48-52
3. Sun, L.S., Wang, S., Shen, M., Beadle, P.J.: A Method for Parameter Estimation of Time-Dependent AR Model Using Higher-Order Spectra and Wavelet Basis. Proc. IEEE Int. Conf. Signal Processing, Vol. 1, (2004) 288-291
4. Fekri, F., Sartipi, M., Mersereau, R.M., Schafer, R.W.: Convolutional Codes Using Finite-Field Wavelets: Time-Varying Codes and More. IEEE Trans. Signal Processing, Vol. 53, No. 5, (2005) 1881–1896
5. Nayebi, K., Smith, M.J.T., Barnwell, T.P.: Analysis-Synthesis Systems on Time Varying Filter Banks. Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Vol. 4, (1992) 617–620

6. Arrowood, J.L., Smith, M.J.T.: Exact Reconstruction Analysis/Synthesis Filter Banks with Time-Varying Filters. Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Vol. 3, (1993) 233–236
7. Herley, C., Kovacevic, J., Ramachandran, K., Vetterli, M.: Tilings of the Time-Frequency Plane: Construction of Arbitrary Orthogonal Bases and Fast Tiling Algorithm. IEEE Trans. Signal Processing, Vol. 41, No. 12, (1993) 3341–3359
8. Queiroz, R.L., Rao, K.R.: Time-Varying Lapped Transforms and Wavelet Packets. IEEE Trans. Signal Processing, Vol. 41, No. 12, (1993) 3293–3305
9. Xia, X.-G., Suter, B.W.: Construction of Perfect Reconstruction Time-Varying FIR Multirate Filter Banks with Overlaps. Proc. IEEE Int. Symposium on Time-Frequency and Time-Scale Analysis, (1994) 120–123
10. Sodagar, I., Nayebi, K., Barnwell, T.P., Smith, M.J.T.: A Novel Structure for Time-Varying FIR Filter Banks. IEEE Trans. Acoust., Speech, Signal Processing, Vol. 3, (1994) 157–160
11. Herley, C., Vetterli, M.: Orthogonal Time-Varying Filter Banks and Wavelets. IEEE Trans. Signal Processing, Vol. 42, No. 10, (1994) 2650–2663
12. Nayebi, K., Sodagar, I., Barnwell, T.P.: Time-Varying Filter Banks and Wavelets. IEEE Trans. Signal Processing, Vol. 42, No. 11, (1994) 2983–2996
13. Herley, C.: Boundary Filters for Finite Length Signals and Time-Varying Filter Banks. IEEE Trans. Circuits and Systems, Vol. 42, (1995) 102–114
14. Gopinath, R.A., Burrus, C.S.: Factorization Approach to Unitary Time-Varying Filter Bank Trees and Wavelets. IEEE Trans. Signal Processing, Vol. 43, No. 3, (1995) 666–680
15. Mau, J., Valot, J., Minaud, D.: Time-Varying Orthogonal Filter Banks without Transient Filters. IEEE Trans. Acoust., Speech, Signal Processing, Vol. 2, (1995) 1328–1331
16. Schwarz, C., Dasgupta, S., Fu, M.: A Lattice Structure for Perfect Reconstruction Linear Time Varying Filter Banks with All Pass Analysis Banks. IEEE Int. Conf. Acoust., Speech, Signal Processing, Vol. 1, (1997) 39–42
17. Schuller, G.: Time-Varying Filter Banks with Variable System Delay. IEEE Int. Conf. Acoust., Speech, Signal Processing, Vol. 3, (1997) 2469–2472
18. Vaidyanathan, P.P.: Multirate Systems and Filter Banks. Prentice Hall, Englewood Cliffs, New Jersey, (1993)
19. Cohen, A., Daubechies, I., Feauveau, J.C.: Bi-orthogonal Bases of Compactly Supported Wavelets. Commun. Pure Apple. Math., Vol. 45, (1992) 485-560
20. Sweldens, W.: The Lifting Scheme: A Custom-Design Construction of Biorthogonal Wavelets. Appl. Comp. Harmon. Anal., Vol. 3, No. 2, (1996) 186-200
21. Kirac, A., Vaidyanathan, P.P.: Theory and Design of Optimum FIR Compaction Filters. IEEE Trans. Signal Processing, Vol. 46, No. 4, (1998) 903–919
22. Shark, L.-K., Yu, C., Smith, J.P.: Genetic Construction of Wavelet Filter Banks for Optimal Denoising of Ultrasonic Signals. IEEE Int. Conf. Signal Processing, Vol. 1, (2002) 197-200
23. Wang, H.J., Chen, T., Peng, S.L.: A Novel Method for Designing Adaptive Compaction Orthogonal Wavelet Filter Banks. Proc. IEEE Int. Conf. Image Processing, Vol. 1, (2003) 1041-1044
24. Shi, G.M., Ding, A.L., Jiao, L.C.: A new approach for Constructing Match Wavelet to Signal Detection. IEEE Int. Conf. Communications, Circuits and Systems, Vol. 2, (2004) 738–741

25. Yang, J.-W., Tang, Y.Y., Zhang, W.-P., He, Z.-Y.: Construction of orthogonal Wavelet Filters in Terms of Unitary Transform. Proc. IEEE Int. conf. Machine Learning and Cybernetics, Vol. 6, (2004) 3908–3912
26. Chan, S.C., Pun, C.K.S., Ho, K.L.: New Design and Realization Techniques for a Class of Perfect Reconstruction Two-Channel FIR Filter Banks and Wavelets Bases. IEEE Trans. Signal Processing, Vol. 52, No. 7, (2004) 2135–2141
27. Sersic, D.: Wavelet Filter Banks with Adaptive Number of Zero Moments. Proc. IEEE Int. Conf. Signal Processing, Vol. 1, (2000) 325-328
28. Vaidyanathan, P.P., Hoang, P.-Q.: Lattice Structure for Optimal Design and Robust Implementation of Two-Channel Perfect Reconstruction QMF Banks. IEEE Trans. Acoust., Speech, Signal Processing, Vol. 36, No. 1, (1988) 81–94
29. Oraintara, S., Tran, T.D., Heller, P.N., Nguyen, T.Q.: Lattice Structure for Regularity Linear-Phase Filter Banks and M-band orthogonal wavelets. IEEE Trans. Signal Processing, Vol. 49, No. 11, (2001) 2659–2672

An Assignment Model on Traffic Matrix Estimation

Tang Hong , Fan Tongliang, and Zhao Guogeng

Chongqing University of Post and Telecommunications
Chongqing, 400065, China

Abstract. It is important to acquire accurate knowledge of traffic matrices of networks for many traffic engineering or network management tasks. Direct measurement of the traffic matrices is difficult in large scale operational IP networks. One approach is to estimate the traffic matrices statistically from easily measured data. The performance of the statistical methods is limited due to they rely on the limited information and require large amount of computation, which limits the convergence of such computation. In this paper, we present an alternative approach to traffic matrix estimation. This method uses *assignment model*. The model is based on the link characters and includes a fast algorithm. The algorithm combines statistical and optimized tomography. The algorithm is evaluated by simulation and the simulation results show that our algorithm is robust, fast, flexible, and scalable.

1 Introduction

Many decisions that IP network operators make depend on how the traffic flows in their network. *Traffic Matrix* (TM) reflects the volume of traffic that flows between all possible pairs of origins and destinations in a network. When used together with routing information, the traffic matrix gives the network operator valuable information about the current network state, routing protocol configuration and management of network congestion, etc. It is impossible to set up internal monitors at all routers in the networks due to cost and deployment problems so that directly measuring TMs is very difficult. In another hand such measurement would impose on the regular behavior of the network because of large amounts of data to be collected. The challenge is then to obtain TMs using limited data measured and information obtained from the network.

Previous work on obtaining traffic matrix usually relied on statistical inference methods that use limited measurements to estimate TM. The term *Network Tomography* [3] was coined for this problem when the partial data come from repeated measurements of the traffic flowing along directed links in the network. In order to handle the estimation problem, Vardi [3] assumes a Poisson model for the traffic matrix estimation in which the covariance of the link loads is used as an additional constraint condition. The traffic matrix is estimated through Maximum Likelihood (ML) estimation. Cao et al. [4] proposed a modified EM algorithm to estimate TM in which a more general scaling law between means and variances of traffic is adopted to find maximal likelihood parameters based on Gaussian model and convergence is hastened by using second-order methods. Tebaldi and West [1] presented a solution which is

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 295 – 304, 2006.

also based on the Poisson model and the Bayesian approach. Since posterior distributions are hard to calculate, they use a Markov Chain Monte Carlo simulation to simulate the posterior distribution. It is also possible to formulate the traffic matrix estimation as a constrained optimization problem that could be solved by using the methods such as Linear Programming [5]. Those statistic techniques produce unacceptable error rates to ISPs. Medina et al. [6] made a comparison between [3] and [4] and proposed a novel choice model to estimates the probability of sending a packet from an origin node to a destination node in the network. Zhang et al. [7] introduced a gravity model and assumed that the amount of traffic from or to a given node was only related to this node. The estimation of TM by this model is consistent with measured link loads at the network edge, but not necessarily so in the interior links. Zhang et al. [8] presented an information-theoretic approach with the Kullback-Leibler distance to minimize the mutual information between source and destination. Medina et al.[10] proposed a two-step approach to infer network traffic demands. First, alternative models, for instance choice model [6] or gravity model [7], are evaluated to generate good starting points for iterative statistical inference techniques. Second, the generated starting point is provided to a statistical inference technique. The division of the TM estimation process into two steps offers great flexibility for combining and evaluation different strategies. In all papers mentioned above, the routing is considered to be constant. Nucci et al [11] presented an approach of changing routing and shifting link loads to infer the traffic matrix.

In this paper, we present an assignment model based on the link characteristics and inspired by Transportation and Traffic Theory [12], propose two algebraic operations to analyze the network and a new algorithm *Expectation Error Rectify Algorithm (*EERA), which combines statistical and optimized tomography.

The remaining parts of this paper are organized as follows. Section 2 introduces the problem and notation. Section 3 introduces the estimation model. Section 4 describes the details of EERA. Section 5 gives the simulation results. Finally, Section 6 makes conclusion.

2 Problem Descriptions

The problem of inferring network traffic matrix can be formulated as follows. Consider a network with N nodes and L directed links. Such a network has $P = N\times(N\text{-}1)$ pairs of origin-destination pairs. Although conceptually traffic matrix is represented in a matrix X, with the amount of data transmitted from node n to node m as element x_{nm}, it is more convenient to use a vector representation. Thus, we enumerate all P origin-destination pairs, and let x_p denote the point-to-point demand of node pair p. For simplicity, we will assume that each point-to-point demand is routed on a single path. The paths are represented by a routing matrix R. R is a $\{0, 1\}$ matrix with rows representing the links of the network and columns representing the OD pairs. Element r_{ij}=1 if link i belongs to the path associated to OD pair j, and r_{ij}=0 otherwise. Let $Y=(y_1,\cdots,y_L)$ be the vector of link counts where y_j gives the link count for link j. Then, X and Y are related via:

$$RX = Y \quad \text{or} \quad \sum_{j=1}^{P} r_{ij} x_j = y_i \ . \tag{1}$$

The matrix R can be obtained by gathering the OSPF or IS-IS links weights and computing the shortest-paths between all OD pairs. The link counts Y are available from the SNMP data. The traffic matrix estimation problem is simply the one of estimating the non-negative vector x based on knowledge of R and y. The challenge in this problem comes from the fact that this system of equations tends to be highly underdetermined: there are typically many more origin-destination pairs than links in a network, and the formula (1) has many more unknowns than equations. So some sort of side information or assumptions must then be added to make the estimation problem well-posed.

3 Assignment Model

The traffic matrix is necessary for many network planning functions. It is clearly necessary to know the volume of expected demand in order to plan the network adequately to handle that traffic with satisfactory quality (low delay and loss of transmitted data). Traffic on any link of the network depends on the way traffic is routed. In current IP networks, the path of a given origin-destination pair is the shortest one, in which the "length" of a link is an administratively assigned weight. Routing for a given origin-destination demand is thus fixed. It is therefore very important to plan route assignments carefully to avoid the demand on any link from overload. Our proposed assignment model is one of such solutions.

The traffic matrix estimation is equivalent to finding a reasonable OD matrix X which reproduces count data of the observed links when X is assigned to the network. In a practical application the reproduction might not be exactly achieved for all traffic counts because of different representation of traffic collection at different times or in the different aggregated network. Therefore, we believe that it is crucial to use the assignment technique in OD matrix estimation based on link counts.

3.1 Link Assignment Coefficient

The link counts are the sum of the OD pairs that are routed across that link. If a link is shared by a large number of OD pairs, then it may be hard to disambiguate how much bandwidth belongs to each OD pair. Each value of OD pair is different proportion out of the link load. We define *assignment coefficient* l_{kz} and let l_{kz} denote the fraction that OD pair x_z is in the link load y_k. Thus, the traffic of OD pair x_z can be described by: $x_z = l_{kz}\, y_k$.

$$y_k = \sum_{z=1}^{p} r_{kz} x_z \quad \text{and} \quad \sum_{z} l_{kz} = 1 \ . \tag{2}$$

Thus, T values can be got from each OD pair x_z ($0 \le T \le L$).

Assignment coefficient usually depends on the policies of traffic exogenous or endogenous determination in the network, for example, congestion in the network and

route characteristics. Assuming independence between the traffic volumes and proportion $l_{kz,}$ the link counts are proportional to the OD volume. In order to reduce the computational complexity of the OD matrix estimation model, we define another coefficient to dispose the data.

3.2 Weight Coefficient

It is true that some OD pairs are more difficult to estimate than others. It would be interesting to investigate the properties of the paths associated with such "troublesome" OD pairs. We assign some weights to each OD pair depend on links, so that we can dispose different values of the OD pair. Let x_{zt} denote one of the values of OD pair x_z. Note that $t \in T$. Each of x_{zt} had different weight during estimating x_z. So we introduce a new coefficient — *weight coefficient*, w_{zt}. The ultimate value of OD pair x_z can be described by:

$$x_z = \sum_t w_{zt} x_{zt} \quad \text{and} \quad \sum_t w_{zt} = 1 \quad . \tag{3}$$

That is, estimate volume of OD pair x_z is expectation of all of x_{zt}.

The values of l_{kz} and w_{zt} can be determined before estimation of the OD matrix is done and taken as exogenous given. The "all-or-nothing" assignment method can be used: all-or-nothing assignment of traffic is obtained when all traffic, for all OD pairs, is assigned to the cost minimizing path. Or, the equilibrium assignment is also a more realistic approach. We have made use of Wardrop's first equilibrium principle in our model. It is that the traffic is in "equilibrium" when no POP can achieve a lower cost by switching to another POP.

Once we have specified the value of assignment coefficient l_{kz} and weight coefficient w_{zt}, the estimation problem is reduced to the combine of constraint program problem and statistical problem. We named it as *assignment model*. The use of assignment model was motivated by wonder which OD pair would influence the load through the link and which OD pair is difficult to estimate. We introduce assignment coefficient and weight coefficient, because we think that the two coefficients are much more stable than the OD matrix itself during the measurement period. What's more, the additional data or assumption about the travel behavior can be used to find a unique traffic matrix.

3.3 Algebraic Operation

The routing matrix summarizes the network traffic structure in useful way. In order to realize the network framework and its characteristic, we present two types multiplication operation to the routing matrix R.

First: $$S = R \times R^T . \tag{4}$$

R multiplied by R^T (matrix transpose) is an L by L matrix $S_{L \times L}$. Entries s_{ij} of matrix S have different signification. While $i = j$, s_{ij} is the total of OD pair through the link i or j, otherwise, s_{ij} is the total of OD pair through the link both i and j.

Second, $$Q = R^T \times R. \tag{5}$$

R^T multiplied by R is a P by P matrix $Q_{P\times P}$. Element q_{ij} of matrix Q has also different meaning. While $i = j$, q_{ij} is the path length of the OD pair i or j. While $i \neq j$, q_{ij} is the sum of the link counts that is shared by both OD pair i and OD pair j.

Applying the two operations, we can obtain the correlation between the OD pairs and find the vital link that impacts the accuracy of estimated method.

4 Algorithm

Any method must provide a solution to the traffic matrix estimation from limited data. This indicates that a new approach incorporating additional knowledge about the network is needed. Indeed non-statistical knowledge about how networks are designed is available to network operators. It would achieve more accurate traffic matrix estimation if combined with statistical data.

The goal of our algorithm is to find the optimization of the following equation:

$$\text{minimize } \left\| RX - Y \right\|_2^2 . \tag{6}$$

$$\text{sub} \quad x_{ij} \geq 0 . \tag{7}$$

That is, the distance between the estimated OD matrix and the real OD matrix is minimized subject to the observed link counts.

The algorithm contains three nested iterative processes. The first is to compute the OD pairs based on the link counts. The second is to calculate the expectation of OD pairs. The third is to validate the constrained term and adjust the traffic matrix. When the total measured and total estimated link loads in each direction converge, this reckoned process ends.

The pseudo-code of our algorithm is listed following:

① Network topology, link vector Y and error threshold e are inputs;
② Computer shortest paths for all OD pairs. Using the operation described in the section 3.3, find the "vital" link and the "troublesome" OD pairs;
③ Iterate as follow to determine the OD pairs.
   (Ⅰ) compute OD pair by: $x_{zk} = l_{kz}\, y_k$;
   (Ⅱ) calculate expectation of OD pairs by: $x_z = \sum_t w_{zt} x_{zt}$;
   (Ⅲ) test the acquired OD pair by: $RX = Y'$.
④ Let $\varepsilon = Y - Y'$ denotes the error vector. If $\varepsilon < e$, the algorithm end. Otherwise, make ε an relaxation variable turn to the step ③.

The results estimated must satisfy equation (1),. Therefore it is vital that there is no error in measurement of the link loads. If the link count data are wrong, it is impossible to get accurate traffic matrix by any algorithm based on link count data. Our approach is not to incorporate additional constraint, but rather to use assignment model to obtain initial estimate, which needs to rectify to satisfy the constraint.

5 Performance

It is not possible to obtain an entire "real" traffic matrix via direct measurement. Therefore assessing the quality of TM estimations and validating TM models is difficult because one cannot compare an estimated TM to "the real thing". There would be no inference problem if real TMs could be obtained.

In order to evaluate TM estimation methods, we need the data about routing, traffic matrix elements and link loads to be consistent. We get the data by the simulate tool OPNET. Our algorithm assumes that Dijkstra's Minimum Weight Path Algorithm is used to the traffic routing. We also assume that all nodes can originate and terminate traffic.

5.1 Topology

We consider a small scale network with 4 nodes topology, as depicted in Figure 1.

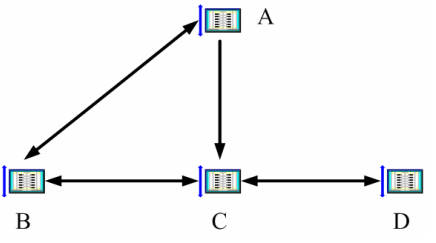


Fig. 1. 4-Node Topology

This network has $L = 7$ directed links and $P = 12$ OD pairs; the ordered sequence of nodes comprising these links and OD pairs appears in the flowing. The corresponding 7×12 routing matrix R is:

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

We use this simple case because it allows us to enumerate all the OD pairs and link counts, which is useful for illustrating how the method behaves. It had also been used in many pioneer papers [3], [6].

observed link: y_1 : A→B, y_2 : B→A, y_3 : A→C, y_4 : B→C, y_5 : C→B, y_6 : C→D, y_7 : D→C;

OD pair: x_1 : A→B, x_2 : A→C , x_3 : A→C→D, x_4 : B→A , x_5 : B→C , x_6 : B→C→D, x_7 : C→B→A, x_8 : C→B, x_9 : C→D, x_{10} : D→C→B→A, x_{11} : D→C→B, x_{12} : D→C .

5.2 Comparison of Algorithms

In this section, we evaluate our method and three methods of others. The optimization approach [2] presents a linear program (LP). VPossion and CGauss separately described in [3] and [4]. The evaluation results of the four methods are presented in Table 1.

Table 1. The compared result among four algorithms

OTM	LP		VPoisson		EERA		CGauss		
	ETM	E%	ETM	E%	ETM	E%	OTM	ETM	E%
AB:318	318	0	318	0	318	0.0	318.65	318.65	0
AC:289	601	107	342	18	297	2.8	329.48	286.98	13
AD:312	0	100	259	17	304	2.6	277.18	318.36	15
BA:294	579	96	334	14	300	2.0	298.14	298.14	0
BC:292	559	91	310	6	271	7.2	354.81	360.97	1.6
BD:267	0	100	249	7	288	7.9	355.39	347.94	2
CA:305	303	0.6	291	5	293	3.9	327.20	317.34	3
CB:289	0	100	361	25	290	0.3	330.04	373.65	13
CD:324	903	178	395	22	311	4.0	253.01	217.32	14
DA:283	0	100	257	9	289	2.1	320.50	329.07	3
DB:277	851	207	245	12	282	1.8	291.52	246.60	15
DC:291	0	100	349	20	280	3.1	310.40	344.82	11

The table 1 shows the original traffic matrix (OTM), the estimated value (ETM) for each OD pair and the relative error (E%). The average error was 98% for the LP method, 13% for the VPoisson method, 3.2% for the EERA method and 7.6% for the CGauss method. The LP method clearly is the worst in the four methods. Although this approach may have worked in the 3-node topology considered in [2], it seems not seem to be effective in the classical 4-node topology. The reason can be that LPs are indeed sensitive to the topology. The data show that EERA is the best method in terms of both the average error and the biggest error. What's more important, our method computation complexity is proportion to n^3, however, other approaches computation complexity is proportion to n^5.

5.3 Sensitivity

We have compared these techniques with respect to the estimation errors yielded. Next we will analyze the sensitivity to prior information and modeling assumptions on OD pairs of our algorithm.

We use synthetic traffic matrices since real TMs are not available. The best TM subjected to certain random model can be obtained by generating synthetic TMs based on what the properties of real TMs are. We need also to generate synthetic TMs that exhibit certain properties that expose the strengths and the weaknesses of the evaluated techniques. With this goal in mind, we generate six types of synthetic TMs that differ in the distribution used to generate their elements. Specifically, we consider *constant*, *Exponential*, *Poisson*, *Gaussian*, *Uniform* and *Mixture* TMs. Up to now,

human being have not discovered the random model of OD pair distributions. We suspect that these are no distributions that may properly reflect the distributions of OD traffic demands in the Internet backbone. So, we design the mixture TMs, which each node have different attribute of delivering packets. We consider the Uniform scenario because this is what is often used by researchers who need a traffic matrix in evaluating algorithms. We include the Constant case because this should be the easiest for these methods to estimate. To compare with [3], [4], we also include poisson and gauss case. All of scenario can be obtained by assigning a attribute of node, constant(1.0), uniform(0.5,1.5) , poisson (1.0), exponential (1.0), normal(1.0,0.2).

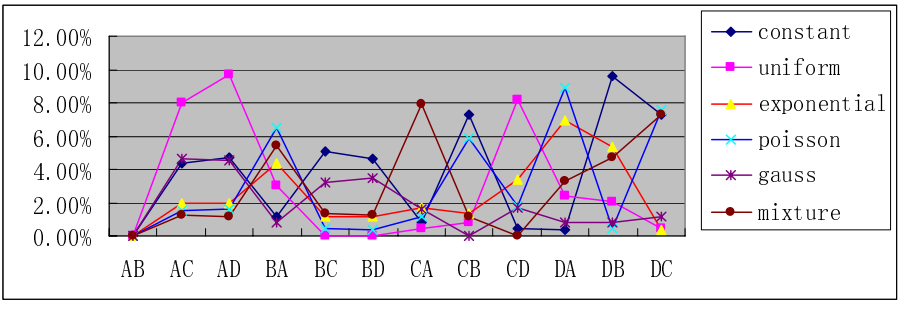


Fig. 2. A comparison of the error to the traffic matrix elements estimated by our algorithm. The curve lines show the error of different scenarios.

Figure 2 shows the error of each OD pairs in different scenario. Obviously, the error is related to the prior information. That is, the maximum error of OD pair is different in the different case. For example, the AD pair is the worst in the Uniform scenario; however, the DB pair is the worst in the Constant scenario. From the Figure 2 we can think that estimation errors may be correlated to heavily shared links, because the assignment model is based on the links. For a given traffic matrix, there are always some OD pairs that can be estimated very closely, while at the same time, other OD pairs are estimated very poorly. What's more, the assumption of independence between the OD pairs can be inaccurate. So, we think that it will be a new approach to find the correlation between the OD pairs. In order to evaluate our method sensitivity to prior information, we show the average error in Figure 3.

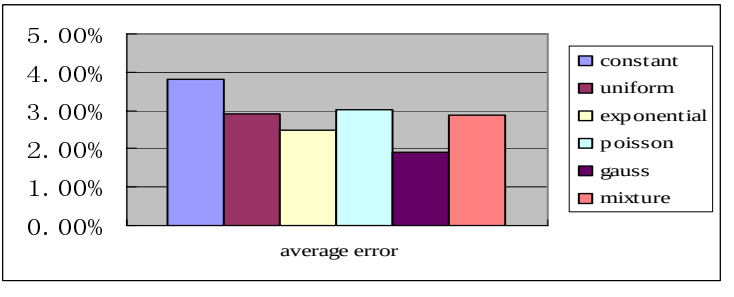


Fig. 3. The average error of different scenarios

Figure 3 shows the results of our method. These results indicate that our method is not sensitive to the distribution, and is more robust to various types of OD pair distributions. From the both figure, we can get a result that the "troublesome" OD pairs are related to not only topology configure but also "real" traffic.

6 Conclusion and Future Work

Availability of an O-D traffic matrix is essential if network "what if" analyses are desired. Estimation techniques based on partial information are used to populate traffic matrix because amassing sufficient data from direct measurements to populate a traffic matrix is typically prohibitively expensive. To handle the problem, we consider the links attribute. For example, how is the link loads aggregated by OD pairs that are routed through the link.

We have presented a new model and introduce a new algorithm. We have validated the method through simulate tool. The results show remarkable accuracy in the small scale network. Our study leaves many important issues unexplored. For example, the data set does not contain measurement errors or component failures and we have not evaluated the effect of such events on the estimation. We have not considered how sensitive traffic engineering tasks are to estimation errors in different demands, and how such information could be incorporated in the estimation procedures. We will expand our method to large network. Another interesting topic for future work would be to understand the nature of the worst-case bounds, and see if they could be exploited in other ways.

Acknowledgement

This paper is supported by *Chunhui* project funded by Ministry of Education, Nature Science Foundation of Chongqing(2005BB2067) and 4G-research special fund of CQUPT.

References

1. Tebaldi. C., West. M., Bayesian Inference of Network Traffic Using Link Count Data. J. of the American Statistical Association, June (1998)557–573
2. Feldmann. A., Greenberg. A., Lund. C., Reingold. N., Rexford. J., True. F., Deriving Traffic Demand for Operational IP Networks: Methodology and Experience. Proceedings of ACM SIGSOMM 2000, Computer Communication Review, Vol. 30, No. 4(2000)
3. Vardi. Y., Network Tomography: Estimating Source-Destination Traffic Intensities from Link Data. J. of the American Statistical Association,(1996)365–377
4. Cao. J., Davis. D., Vander Wiel. S., Yu, B., Time-Varying Network Tomography: Router Link Data Journal of the American Statistical Association, Vol 95, No 452, (2000)1063-1075.
5. Goldschmidt. O., ISP Backbone Traffic Inference Methods to Support Traffic Engineering. In Internet Statistics and Metrics Analysis (ISMA) Workshop, San Diego, CA, December (2000)

6. Medina. A., Taft. N., Salamatian. K., Bhattacharyya. S., Diot. C. Traffic Matrix Estimation: Existing Techniques Compared and New Directions. In ACM SIGCOMM, Pitsburgh, PA,(2002)
7. Zhang. Y., Roughan. M., Duffield. N. and Greenberg. A.. Fast Accurate Computation of Large-Scale IP Traffic Matrices from Link Loads. ACM SIGMETRICS,(2003)
8. Zhang. Y., Roughan. M., Lund. C., David Donoho. An Information-Theoretic Approach to Traffic Matrix Estimation. ACM SIGCOMM, August (2003)
9. Abrahamsson. T., Estimation of Origin-Destination Matrices using Traffic Counts-A. Literature Survey. Technical Report IR-98021, International Institute for Applied Systems. Analysis, (1998)
10. Medina. A., Salamatian. K., Taft. N., Matta. I., Diot. C., A Two-step Statistical Approach for Inferring Network - Traffic Demands Medina (2004) www.cs.bu.edu/ techreports/ps/2004-011-two-step-tm-inference.ps
11. Nucci. A., Cruz. R., Taft. N., Diot. C., Design of IGP link weight changes for estimation of traffic matrices, in Proc. IEEE INFOCOM, Hong Kong, March (2004)
12. Bell. M.G.H., Lan. W.H.K., Ploss. G., Inaudi. D., Stochastic User Equilibrium Assignment and Iterative Balancing, Transportation and Traffic Theory/C,F.D. aganzo(Editor), Elsevier Science Publishers, (1993)427-440

M-Channel Nonuniform Filter Banks with Arbitrary Scaling Factors*

Xuemei Xie, Liangjun Wang, and Siqi Shi

School of Electronic Engineering, Xidian University, Xi'an, P.R. China
xmxie@see.xidian.edu.cn,
{lj_wang, sqshi}@mail.xidian.edu.cn

Abstract. In conventional filter banks, the sampling factors are restricted to rational numbers and frequency partition is always rather inflexible, stemming from the fact that certain constraint on each subband position is always placed. In this paper, we present a class of *M*-channel nonuniform filter banks with arbitrary sampling factors including integer, rational, and even irrational numbers. Consequently, the frequency partitioning in the proposed filter bank is much more flexible, which is very attractive in many applications.

1 Introduction

Filter banks have been widely used in many applications such as subband coding due to its quality in time-frequency decomposition. The filter banks with only two subbands were proposed in the 70s to 80s of last century. Later, *M*-channel filter banks appeared, which allow us to partition the signal more precisely. In these filter banks, the bandwidths of all subbands are equal and therefore they are called uniform filter banks. However, in many applications such as speech coding, the energy distribution of the signal is not uniform. Filter banks which can split the signal into unequal parts, known as nonuniform filter banks, are therefore much desired, due to their flexibility in frequency partitioning. A great deal of the work has been done on the nonuniform filter banks [1-7]. The integer-decimated filter banks [1] were studied at early stages, in which the sampling factors are all integers and the subband bandwidth have the form of π / p_k, where p_k are integers. Then, more flexible nonuniform filter banks [2-7] with rational decimation factors were brought forward, in which the sampling factors are in the form of p_k / q_k, where p_k and q_k are mutually prime. A primary drawback of these systems is that there exist some constraints on the position of the subbands, which makes its application very limited. Although this problem can be solved by shifting the frequency band of each channel to a suitable band before the decimation [4-5], the system seems very complicated, because in each channel we need to multiply the filtered signal with two sequences to move the band to a proper position. Furthermore, the sampling factors in these systems are still restricted to

* Work supported by NSF of China (Grant No. 60372047) and NSF of Shaanxi Province, China (Grant No. 2005F18).

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 305 – 314, 2006.

rational numbers. When the sampling factors are irrational, the structures of the systems described above do not work because any irrational number cannot be expressed exactly as a fraction.

Some early works on the irrational scaling operation were found in [8-10]. In [11], based on the idea of the frequency warping [12-14], a concise closed formula in time domain was proposed. Different from the traditional decimation, the formula combines an ideal low pass filter with a decimation operation. However, only two-channel nonuniform filter bank was achieved.

In this paper, we will extend the traditional decimation/interpolation operation to a more general case, in which the decimation factor can even be an irrational number. Based on this concept, we construct an M -channel filter bank with arbitrary sampling factors. In this system, the first and the last channels are easy to realize. However, as for the other channels, it is much more involved to design them. Basically, there exist two main problems: a) how to extract the frequency; and b) how to decimate with no aliasing. Some appropriate schemes are proposed in this paper, aiming at solving those problems. In particular, we propose a novel structure to deal with the second problem by dividing the decimating operation into two blocks and putting a modulator $(-1)^n$ between them. This new filter structure enables the filter banks to be realized with an arbitrary frequency partitioning.

The organization of the paper is as follows, Section 2 presents the arbitrary scaling formula. Section 3 discusses the structure of the proposed system. The simulation result is given in Section 4. Conclusions are drawn in Section 5.

2 The Arbitrary Scaling Operations

Traditionally, the decimation and interpolation operations refer to the integer sampling factors [15, 16]. In this section, we will extend these two concepts to more general cases.

2.1 Arbitrary Decimation

Suppose there is a discrete time signal $x(n)$, and we denote its DTFT by $X(e^{j\omega})$. Decimating the signal by a, where a is an integer and $a>1$, is easy to be realized since we can preserve the signal at the time an, $n\in\mathbb{N}$. However, if a is a rational or irrational number, it becomes invalid.

Now we will show the fact that no matter what a is, integer, rational or irrational number, decimating the signal $x(n)$ by a can be realized if the following procedures are taken:

(a) From the digital signal $x(n)$, the corresponding continuous time signal $x(t)$ can be obtained. The relation between $x(n)$ and $x(t)$ is that the signal $x(n)$ is sampled from $x(t)$ with the time period T. For a good expression, we need to define a function $X'(e^{j\omega})$,

$$X'(e^{j\omega}) = \begin{cases} X(e^{j\omega}), & -\pi \le \omega \le \pi, \\ 0, & \text{others.} \end{cases} \tag{1}$$

If we denote continuous time Fourier transform of $x(t)$ by $X(j\Omega)$, then with $\omega = \Omega T$ we have

$$X(j\Omega) = TX'(e^{j\Omega T}). \tag{2}$$

(b) Sample $x(t)$ with new time period aT, the sampled signal is what we want finally. Here, we denote it by $x_a(n)$ and its DTFT $X_a(e^{j\omega})$, and

$$X_a(e^{j\omega}) = \frac{1}{aT}\sum_{k=-\infty}^{\infty} X\left(j\left(\frac{\omega}{aT} + \frac{2k\pi}{aT}\right)\right) = \frac{1}{a}\sum_{k=-\infty}^{\infty} X'\left(e^{j\left(\frac{\omega}{a}+\frac{2k\pi}{a}\right)}\right). \tag{3}$$

According to the definition of IDTFT, we can get,

$$\begin{aligned} x_a(n) &= \mathrm{IDFT}\left\{X_a(e^{j\omega})\right\} = \frac{1}{2\pi a}\int_{-\pi}^{\pi}\sum_{k=-\infty}^{\infty} X'\left(e^{j\left(\frac{\omega}{a}+\frac{2k\pi}{a}\right)}\right)e^{j\omega n}d\omega \\ &= \frac{1}{2\pi a}\int_{-a\pi}^{a\pi} X(e^{j\omega/a})e^{j\omega n}d\omega. \end{aligned} \tag{4}$$

Eq. (4) shows the relationship in frequency domain between the original discrete signal $x(n)$ and the decimated signal $x_a(n)$.

For the purpose of the later use, we would like to discuss briefly the implication of the above results in the context of the filter bank. Basically, a filter bank consists of a set of filters with different subbands performing frequency partitioning. In general, a decimation operation is always done after filtering, giving rise to a lower sampling rate.

Assume the filter is an ideal lowpass one with the passband being $[-\pi/b, \pi/b]$, the decimation factor is a, and the signal after the filtering and decimation operation is $x_{b,a}(n)$. The signal $x_{b,a}(n)$ can be derived readily from Eq. (4) by changing the integration range from $[-a\pi, a\pi]$ to $[-a\pi/b, a\pi/b]$. More specifically, we have

$$\begin{aligned} x_{b,a}(n) &= \frac{1}{2\pi a}\int_{-\frac{a\pi}{b}}^{\frac{a\pi}{b}} X(e^{j\omega/a})e^{j\omega n}d\omega = \frac{1}{2\pi a}\sum_{k=-\infty}^{+\infty} x(k)\int_{-\frac{a\pi}{b}}^{\frac{a\pi}{b}} e^{j(n-k/a)\omega}d\omega \\ &= \frac{1}{2\pi a}\sum_{k=-\infty}^{+\infty} x(k)\frac{2\pi a}{b}\mathrm{sinc}\left(\frac{an}{b}-\frac{k}{b}\right) \\ &= \frac{1}{b}\sum_{k=-\infty}^{+\infty} x(k)\mathrm{sinc}\left(\frac{an}{b}-\frac{k}{b}\right). \end{aligned} \tag{5}$$

When $b > a$, the sampling is called oversampling. And when $b = a$, it is critical sampling. Then, Eq. (5) can be written as (6)

$$x_{a,a}(n) = \frac{1}{a}\sum_{k=-\infty}^{+\infty} x(k)\mathrm{sinc}\left(n - \frac{k}{a}\right). \tag{6}$$

When $b = 1$, which means that the signal is decimated directly without any filter. In this case, Eq. (5) becomes

$$x_{1,a}(n) = x_a(n) = \sum_{k=-\infty}^{+\infty} x(k)\mathrm{sinc}(an - k). \tag{7}$$

2.2 Arbitrary Interpolation

Similar to the derivation above, the signal $x(n)$ is interpolated by a factor a, $a \in (1, +\infty)$. The DTFT of the resulting signal $x_a(n)$ is defined as follows,

$$X_a(e^{j\omega}) = X(e^{ja\omega}), \tag{8}$$

or in time domain,

$$x_a(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(e^{ja\omega})e^{j\omega n}d\omega. \tag{9}$$

In filter banks, after interpolation the low frequency part $(-\pi/b, \pi/b)$ of the interpolated signal is preserved by a filter in the synthesis part, the output after filtering denoted by $x_{a,b}(n)$ can be written as

$$\begin{aligned} x_{a,b}(n) &= \frac{a}{2\pi}\int_{-\frac{\pi}{b}}^{\frac{\pi}{b}} X(e^{j\omega a})e^{j\omega n}d\omega = \frac{a}{2\pi}\sum_{k=-\infty}^{+\infty} x(k)\int_{-\frac{\pi}{b}}^{\frac{\pi}{b}} e^{j(n-ka)\omega}d\omega \\ &= \frac{a}{2\pi}\sum_{k=-\infty}^{+\infty} x(k)\frac{2\pi}{b}\mathrm{sinc}\left(\frac{an}{b} - \frac{k}{b}\right) \\ &= \frac{a}{b}\sum_{k=-\infty}^{+\infty} x(k)\mathrm{sinc}\left(\frac{n}{b} - \frac{ak}{b}\right). \end{aligned} \tag{10}$$

When $b = a$, substituting $b = a$ into Eq. (10), then (10) becomes

$$x_{a,a}(n) = \sum_{k=-\infty}^{+\infty} x(k)\mathrm{sinc}\left(\frac{n}{a} - k\right). \tag{11}$$

If $b = 1$, which corresponds the case of no filter used. Rewriting Eq. (10) by substituting b with 1, then we get Eq. (12),

$$x_{a,1}(n) = x_a(n) = a\sum_{k=-\infty}^{+\infty} x(k)\mathrm{sinc}(n-ak). \tag{12}$$

To summarize, we have proposed the concepts of arbitrary decimation and interpolation, which are the extension of the traditional concepts. The realizations of these two operations are given in Eq. (7) and (12). In addition, we combine the decimation and interpolation with the ideal filtering operation, which are realized by using Eq. (5) and Eq. (10), respectively. All these results will be used in the next section.

It can be easily proved that when a is an integer number, the output signal $x_a(n)$ expressed by Eq. (7) or (12) is the same as the output of traditional decimation or interpolation.

3 *M*-Channel Nonuniform Filter Bank with Arbitrary Scaling Factors

Based on the two concepts given above, a two channel nonuniform filter bank with arbitrary scaling factors was proposed in [11]. The frequency partitioning in this filter bank structure is very flexible. However, it only deals with the two-channel cases. Here, we will try to extent it to *M*-channel cases.

As shown in Fig. 1, the input signal is decomposed into M subbands. In each channel, the corresponding frequency part of the signal should be dealt with. From low to high in frequency, we denote them by channel 0 to channel $M-1$, which covers the regions $(-\pi/a_0, \pi/a_0), (\pi/a_0, \pi/a_1)\cup(-\pi/a_1, -\pi/a_0), ...,$ $(\pi/a_{M-2}, \pi)\cup(-\pi, -\pi/a_{M-2})$, where $a_0 > a_1 > \cdots > a_{M-2} > 1$. The bandwidths of those subbands are $2\pi/a_0$, $2(a_1-a_0)\pi/a_0a_1$, $\cdots$, $2(a_{M-2}-1)\pi/a_{M-2}$ respectively.

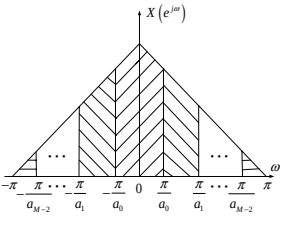


Fig. 1. *M*-channel frequency partition

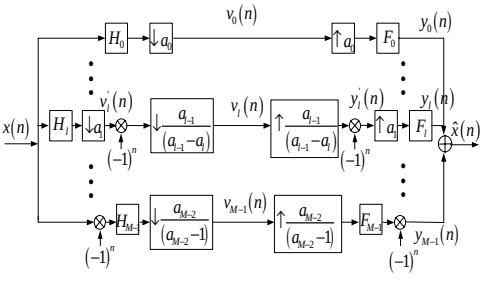


Fig. 2. The system structure

From the frequency partitioning described above, we get the structure of the whole system shown in Fig. 2. In order to realize the frequency partitioning shown in Fig. 1, the filters in channel 1 to channel *M*-2 are all treated as ideal filters with the pass-bands being in the frequency region of the corresponding channel in Fig. 1. To be more specific, the filter H_0 are lowpass ones, $H_1, H_2, \cdots H_{M-2}$ are bandpass. For the

last channel, due to the effect of the modulator $(-1)^n$, the filter in this channel is also an ideal lowpass filter.

The whole filter bank structure is realized in three parts which are divided according to the corresponding channels of the filters. H_0 is a lowpass filter. H_{M-1} is also a lowpass filter. As for the others, filters $H_1, H_2, \cdots H_{M-2}$ are achieved by the manipulation of two ideal lowpass filters. In what follows, we will discuss the realization of the whole system structure in details. The main analysis will be focused on last part.

The first channel: This channel only deals with the low frequency part $(-\pi/a_0, \pi/a_0)$ of the original signal. Thus we can apply an ideal lowpass filter with the pass band being $(-\pi/a_0, \pi/a_0)$ and take a decimation by a_0. Those two operations can be realized in one step directly with the help of (6),

$$v_0(n) = \frac{1}{a_0} \sum_{k=-\infty}^{+\infty} x(k) \mathrm{sinc}\left(n - \frac{k}{a_0} \right). \tag{13}$$

In the synthesis part, (11) is used

$$y_0(n) = \sum_{k=-\infty}^{+\infty} v_0(k) \mathrm{sinc}\left(\frac{n}{a_0} - k \right). \tag{14}$$

The last channel: We must notice that the formulas in Section 2 are derived based on the ideal lowpass filter. Here, it deals with the high frequency part of signal. Therefore the signal has to be shifted by π in frequency domain. This can be done by multiplying the signal with the sequence $(-1)^n$ in time domain. The following operation is the same as channel 0, only need to change the passband of ideal lowpass filter to $(-(a_{M-2}-1)\pi/a_{M-2}, (a_{M-2}-1)\pi/a_{M-2})$, and the decimation factor to $a_{M-2}/(a_{M-2}-1)$. So according to (6), the operation in the analysis part can be realized as follows,

$$v_{M-1}(n) = \frac{1}{a_{M-2}/(a_{M-2}-1)} \sum_{k=-\infty}^{+\infty} (-1)^k x(k) \mathrm{sinc}\left(n - \frac{k}{a_{M-2}/(a_{M-2}-1)} \right), \tag{15}$$

and in the synthesis part, they need to be shifted back. The results can be expressed as follows,

$$y_{M-1}(n) = \sum_{k=-\infty}^{+\infty} (-1)^k v_{M-1}(k) \mathrm{sinc}\left(\frac{n}{a_{M-2}/(a_{M-2}-1)} - k \right). \tag{16}$$

The remaining channels: As described above, there exist two problems with these channels. The first one is how to design filters to obtain the corresponding subband

signals of those channels; And the second one is that even if the needed parts of the original signal are filtered out, they cannot be decimated to full frequency band directly because it is very likely to cause aliasing without being able to recover the signal perfectly in the synthesis part. The details of this discussion can be found in [3, 4].

For the sake of easy discussion, we denote the channel number by $l, l \in 1,2,\cdots M-2$, as shown in Fig. 2. To solve the first problem mentioned above, two ideal lowpass filters with the passband being $(-\pi/a_l, \pi/a_l)$ and $(-\pi/a_{l-1}, \pi/a_{l-1})$ can be used to realize the frequency extraction. By a simple operation of subtracting, an identical bandpass filter with the passband being $(-\pi/a_l, -\pi/a_{l-1}) \bigcup (\pi/a_{l-1}, \pi/a_l)$ is formed.

For the second problem, we proposed a simple and effect approach to solve it. After passing the equivalent bandpass filter, the signal is decimated by a_l firstly other than $a_{l-1}a_l/(a_{l-1}-a_l)$. Because $a_{l-1} > a_l$, there will be no aliasing caused in this operation. Then the output signal is shifted by π and decimated by $a_{l-1}/(a_{l-1}-a_l)$. It is obvious that there will be no aliasing in this step either. The problem is solved by dividing one decimation operation into two steps.

We combine the ideal bandpass filter and the first decimation shown in Fig. 2. The relation between output $v_l'(n)$ and the original signal $x(n)$ is got by Eqs. (5) and (6)

$$\begin{aligned} v_l'(n) &= \frac{1}{a_l}\sum_{k=-\infty}^{+\infty} x(k)\mathrm{sinc}\left(n-\frac{k}{a_l}\right) - \frac{1}{a_{l-1}}\sum_{k=-\infty}^{+\infty} x(k)\mathrm{sinc}\left(\frac{a_l n}{a_{l-1}} - \frac{k}{a_{l-1}}\right) \\ &= \sum_{k=-\infty}^{+\infty} x(k)\left(\frac{1}{a_l}\mathrm{sinc}\left(n-\frac{k}{a_l}\right) - \frac{1}{a_{l-1}}\mathrm{sinc}\left(\frac{a_l n}{a_{l-1}} - \frac{k}{a_{l-1}}\right)\right). \end{aligned} \tag{17}$$

The signal $v_l'(n)$ is then modulated by the sequence $(-1)^n$ and decimated by $a_{l-1}/(a_{l-1}-a_l)$. With the help of (7), these two steps can be written as follows,

$$v_l(n) = \sum_{k=-\infty}^{+\infty} (-1)^k v_l'(k)\mathrm{sinc}\left(a_{l-1}/(a_{l-1}-a_l) * n - k\right). \tag{18}$$

In the synthesis part, the inverse operation can be expressed as

$$y_l'(n) = a_{l-1}/(a_{l-1}-a_l)\sum_{k=-\infty}^{+\infty} v(k)\mathrm{sinc}(n - a_{l-1}/(a_{l-1}-a_l) * k). \tag{19}$$

Finally, we can obtain $y_l(n)$ by Eq. (11)

$$y_l(n) = \sum_{k=-\infty}^{+\infty} (-1)^k y_l'(k)\left(\mathrm{sinc}\left(\frac{n}{a_l} - k\right) - \frac{a_l}{a_{l-1}}\mathrm{sinc}\left(\frac{n}{a_{l-1}} - \frac{a_l k}{a_{l-1}}\right)\right). \tag{20}$$

4 Design Example

In this section, we will give an example to illustrate the validity of this approach. Due to space limitation, only an example of 3-channel filter bank is shown. It should be noticed that in the formula above, the index k has a range from minus infinity to infinity, so we must take a truncation of k in practical implementations. In this experiment, we set $k=[-256\sim+256]$. The parameters of the filter bank system are given in Table 1. The simulation results are shown in Fig. 3.

Table 1. The parameters of the system

Channel number	0	1	2
a_l	$4/\sqrt{2}$	$\sqrt{3}$	1
Passband of the filter	$\left(-\frac{\pi}{4/\sqrt{2}},\frac{\pi}{4/\sqrt{2}}\right)$	$\left(-\frac{\pi}{\sqrt{3}},-\frac{\pi}{4/\sqrt{2}}\right)\cup\left(\frac{\pi}{4/\sqrt{2}},\frac{\pi}{\sqrt{3}}\right)$	$\left(-\frac{2\pi}{(3+\sqrt{3})},\frac{2\pi}{(3+\sqrt{3})}\right)$
Bandwidth	$\frac{\sqrt{2}\pi}{2}$	$\left(\frac{2\sqrt{3}}{3}-\frac{\sqrt{2}}{2}\right)\pi$	$\left(2-\frac{2\sqrt{3}}{3}\right)\pi$

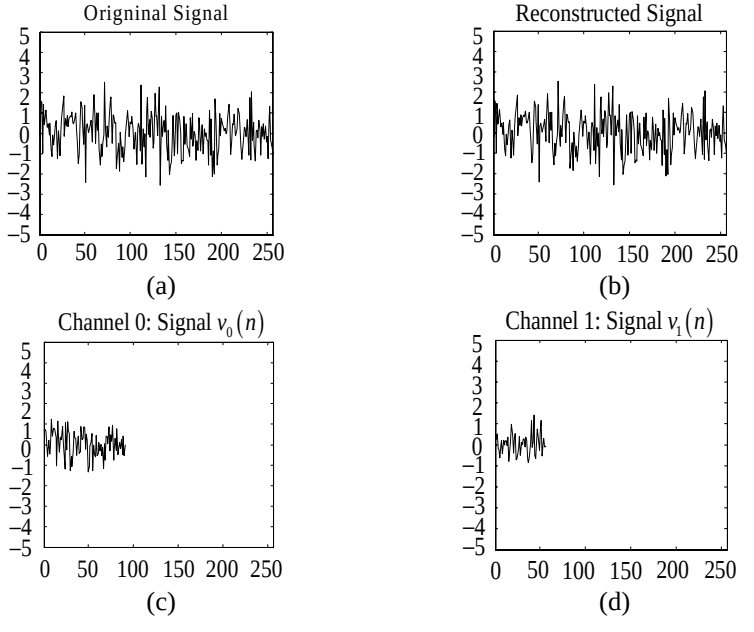
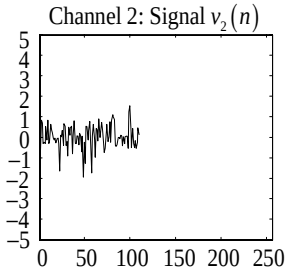
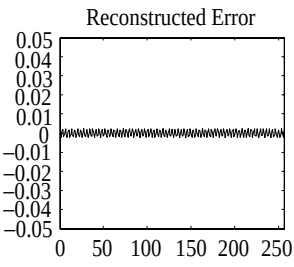


Fig. 3. The experimental results. (a) Original input signal: A 256-point normalized Gaussiaan random signal with zero mean and unit variance; (b) Reconstructed signal; (c), (d) and (e) Subband signals of channel 0, 1, 2, respectively; (f) Reconstructed error.



(e)



(f)

Fig. 3. (*Continued*)

From the experimental results, we can see the reconstruction errors with the mean absolute being about 1.5×10^{-3} is introduced due to the truncation of k, however, which can decrease as k grows larger .

5 Conclusions

A class of *M*-channel nonuniform filter banks with arbitrary scaling factors is proposed. The system casts off the constraint on the each subband in traditional nonuniform filter banks designed by directly structure. So the system we proposed has realized arbitrary frequency partitioning, and which can be well used in many applications.

References

1. Hoang, P.Q., Vaidyanathan, P.P.: Non-uniform multirate filter banks: Theory and design. In Proc. Int. Symp. Circuits Syst. (1989) 371-374
2. Nayebi, K., Barnwell, T.P., Smith, M. J. T.: Nonuniform filter banks: A reconstruction and design theory. IEEE Trans. Signal Processing, Vol. 41. (1993) 1114–1127
3. Kovacevic, J., Vetterli, M.:Perfect reconstruction filter banks with rational sampling factors. IEEE Trans. Signal Processing, Vol. 42. (1993) 2047-2066
4. Li, J., Nguyen, T.Q., Tantaratana, S.: A simple design method for near-perfect-reconstruction nonuniform filter banks. IEEE Trans. Signal Processing, Vol. 45. (1997) 2105-2109
5. Liu, B., Bruton, L.T.: The design of *N*-band nonuniform-band maximally decimated filter banks. in Proc. Asilomar Conf. (1993) 1281-1285
6. Xie, X.M., Shan, S.C., Yuk, T.I.: On the design of a class of PR nonuniform cosine modulated filter banks with flexible rational sampling. IEICE trans. Circuits and Systems 1, Vol.52. (2005) 1965-1981
7. Xie X. M., Chan S. C., and Yuk T. I.: A design of recombination nonuniform filter banks with linear-phase analysis and synthesis filters. IEEE Trans. Signal Processing, accepted for publication, 2006
8. Adams, J.W., Bayma, R.W., Nelson, J.E.: Digital filter design for generalized interpolation. In Proc. IEEE Int. Symp. Circuits Syst., Vol. 2. (1989) 1299-1302

9. Ramstad, T.A.: Digital methods for conversion between arbitrary sampling frequencies. IEEE Trans. Acoust., Speech, Signal Process, Vol. ASSP-32. (1984) 577-591
10. Ramstad, T.A.: Digital two-rate IIR and hybrid IIR/FIR filters for sampling rate conversion. IEEE Trans. Commun., Vol. COM-30. (1982) 1466-1476
11. Pei, S.C., Kao, M.P.: A two-channel nonuniform perfect reconstruction filter bank with irrational down-sampling factors. IEEE Signal Processing Letters, Vol. 12. (2005)
12. Zhao, W., Rao, R.M.: Continuous-dilation discrete-time self-similar signals and linear scale-invariant systems. In Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Vol. 3. (1998) 1549-1552
13. Lee, S., Zhao, W., Narasimha, R., Rao, R.M.: Discrete-time models for statistically self-similar signals. IEEE Trans. Signal Process., Vol. 51. (2003) 1221-1230
14. Zhao, W., Rao, R.M.: Discrete-time continuous-dilation wavelet transforms. In Proc. IEEE-SP Int. Symp. Time-Frequency Time-Scale Anal.. (1998) 233-236
15. Crochiere, R.E., Rabiner, L.R.: Interpolation and decimation of digital signals-a tutorial review. IEEE Proceedings, Vol. 69. (1981) 300-311
16. Vaidyanathan, P.P.: Multirate Systems and Filter Banks. Englewood Cliffs, New Jersey: Prentice Hall. 1992

Variance Minimization Dual Adaptive Control for Stochastic Systems with Unknown Parameters

Gao Zhenbin, Qian Fucai, and Liu Ding

School of Automation and Information Engineering, Xi'an University of Technology, Xi'an, China,710048

Abstract. The variance minimization dual adaptive control approach with unknown parameters for stochastic system is studied in this paper. First,the problem is proposed for solving the optimal dual control law. Furthermore, the differential equation is transferred into state space model, the suboptimal dual control law is obtained through DUL algorithm. Finally, the example is given to verify the method developed in this paper. It is shown that the method is effective and practical.

Keywords: dual control; adaptive control; stochastic systems.

1 Introduction

In 1960, former Soviet Union researcher Feldbaum presented the notion of dual control when he studied the unknown parameter stochastic system[1]. The dual control has the two aims:(i)to drive the system towards a desired state,and (ii) to perform learning to reduce the systems uncertainty. In 2000,IEEE Control Systems Magazine lists the dual control as the one of 25 most prominent subjects in the last century which had significantly impacted the development of control theory. In 1973, Lainiotis et al adopt the adaptive control strategy which base on the weighting of the posterior probability, it is called the DUL algorithm[2]; In 1982, Milito et al lead the innovation with learning coefficient into the performance index, achieving the better compromise between the controlling and the learning, it is famous IDC (Innovations dual control) strategy[3]. In 2002, Li Duan et al present variance minimization method, which have the feature of active learning[4][5]. In 2004, Qian Fu-cai et al present the two level optimal algorithm of dual control law which system model is the differential equation with the unknown parameter[6]. This paper considers the extensive class of dual control problem, which system model is the differential equation and the unknown parameters exist in the definite set which include the finite group of values. A suboptimal dual control law is achieved through the differential equation transferring into the state space model.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 315–318, 2006.

2 Problem Formulation

Consider a time-varying discrete single-input/single-output system described by

$$y(k) = A(q^{-1})y(k) + B(q^{-1})u(k) + w(k) \tag{1}$$

where $A(q^{-1}) = a_1q^{-1} + \cdots + a_nq^{-n}, B(q^{-1}) = b_1q^{-1} + \cdots + b_mq^{-m}, (m \le n)$, $y(k) \in R$,$u(k) \in R$,$w(k) \in R$ are the output, input, and disturbance sequences, respectively. $a_i(i = 1, \cdots, n)$,$b_j(j = 1, \cdots, m)$ are the unknown parameters, q^{-1} is time delay. The noise sequence $w(k)$ is assumed to be Gaussian white noise, that is, $w(k) \sim N(0, R_1)$. Suppose the finite parameters can be select, that is, $\theta = [a_1, \cdots, a_n, b_1, \cdots, b_m], \theta \in \Theta = \{\theta_1, \cdots, \theta_s\}$.

Consider the stochastic optimal problem with the unknown parameter abovementioned

$$\text{(G)} \quad minJ = minE\left\{ y^2(N) + \sum_{k=0}^{N-1} \left[y^2(k) + r(k)u^2(k) \right] \mid I^0 \right\} \tag{2}$$

$$\text{s.t.} \quad y(k+1) = A_i(q^{-1})y(k) + B_i(q^{-1})u(k) + w(k) \tag{3}$$

$$k = 0, 1, \cdots, N-1, i = 1, 2, \cdots s$$

where $I^0 = \{u(-1), \cdots, u(-m-1), y(0), \cdots, y(-n+1)\}$ is the set of original information; $A_i(q^{-1}), B_i(q^{-1})$ are polynomial of the parameter $\theta = \theta_i(i = 1, 2, \cdots, s)$;$r(k) \ge 0(k = 1, 2, \cdots, N-1)$ is weighting coefficient; The priories probability of the parameter θ is $q_i(0) = P(\theta = \theta_i \mid I^0), i = 1, 2, \cdots, s$.

Suppose the information set is I^k at the k stage, $I^k = \{u(0), u(1), \cdots, u(k-1), y(1), y(2), \cdots, y(k), I^0\}$, then, the aim of the dual control is finding control law $u(k) = f_k(I^k), (k = 0, 1, \cdots, N-1)$, which minimize the performance index of problem (G).

3 Suboptimal Dual Control Law

Theorem: The problem (G) of the minimization variance control with the unknown parameter equals to the problem (P)

$$(P) \qquad minJ_1 = minE\{x^T(N)Q_0(N)x(N) + \sum_{k=0}^{N-1} \left[x^T(k)Q_1(k)x(k) + Q_2(k)u^2(k) \right] \mid I^0\} \tag{4}$$

$$s.t. \quad x(k+1) = \Phi_i(k)x(k) + \Gamma_i(k)u(k) + \Lambda_i(k)w(k), i = 0, 1, \cdots, s \tag{5}$$

$$y(k) = Hx(k) + w(k) \tag{6}$$

where $x(k) \in R^n$. When $\theta = \theta_i(i = 1, 2, \cdots, s)$,let $\alpha_i = [\alpha_1^i, \alpha_2^i, \cdots, \alpha_n^i]^T, \beta = [I_{(n-1)}, O_{(n-1)\times 1}]^T$, then $\Phi_i(k) = [\alpha_i, \beta], \Gamma_i(k) = [b_1^i, \cdots, b_m^i, \cdots, b_n^i]^T, (b_t^i = 0, t = m+1, \cdots, n), \Lambda_i(k) = \alpha_i, H = [10 \cdots 0]$, where $a_1^i, \cdots, a_n^i, b_1^i, \cdots, b_m^i, \cdots, b_n^i$ are the parameters of polynomial $A_i(q^{-1}), B_i(q^{-1})$ in the equation (3). $v(k) =$

$\Lambda_i(k)w(k) \sim N(0, R_2(k)), w(k) \sim N(0, R_1).Q_0(N) = Q_1(k) = H^T H, Q_2(k) = r(k), (k = 0, 1, \cdots, N-1)$.

Definition 1: When $\theta = \theta_i$,assuming the state estimation of the k stage is

$$\hat{x}_i(k|k) = E\{x(k) \mid \theta = \theta_i, I^k\}, k = 1, 2, \cdots, N, i = 1, 2, \cdots, s \tag{7}$$

which can be obtained from the Kalman filter.

Definition2: The posterior probability of the model i at the k stage

$$q_i(k) = P(\theta = \theta_i \mid I^k), \ k = 1, \ldots, N, i = 1, \cdots, s \tag{8}$$

posterior probability $q_i(k), i = 1, 2, \ldots, s$ can be calculated from the Bayes formula

$$q_i(k) = \frac{L_i(k)}{\sum_{j=1}^{s} q_j(k-1)L_j(k)} q_i(k-1), \quad k = 1, 2, \ldots, N \tag{9}$$

original condition: $q_i(0)$,where

$$\begin{aligned} L_i(k) &= |P_y(k|k-1, \theta_i)|^{-\frac{1}{2}} \exp[-\frac{1}{2}\tilde{y}^T(k|k-1, \theta_i)P_y^{-1}(k|k-1, \theta_i) \times \\ &\quad \times \tilde{y}(k|k-1, \theta_i)] \\ \tilde{y}(k|k-1, \theta_i) &= y(k) - H\hat{x}_i(k|k-1) \\ P_y(k|k-1, \theta_i) &= HP_i(k|k-1)H^T + R_1. \end{aligned}$$

where $\hat{x}_i(k|k-1)$ is the one-step-ahead predictive estimating values; $P_i(k|k-1)$ is the one-step-ahead covariance on Kalman filter when $\theta = \theta_i, i = 1, 2, \cdots, s$.

According to reference[2],the suboptimal dual control law can be obtained by

$$u(k) = \sum_{i=1}^{s} \hat{u}_i(k|k)q_i(k), (k = 0, 1, \cdots, N-1) \tag{10}$$

$$\hat{u}_i(k|k) = -K_i\hat{x}_i(k|k)(i = 1, 2, \cdots, s)$$

$$K_i(k) = \left[Q_2(k) + \Gamma_i^T(k)S_i(k+1)\Gamma_i(k)\right]^{-1} \Gamma_i^T(k)S_i(k+1)\Phi_i(k) \tag{11}$$

$$\begin{aligned} S_i(k) &= \Phi_i^T(k)S(k+1)\Phi_i(k) + Q_1(k) \\ &\quad -K_i^T(k)\left[Q_2(k) + \Gamma_i^T(k)S_i(k+1)\Gamma_i(k)\right]K_i(k) \end{aligned} \tag{12}$$

boundary condition: $S_i(N) = Q_0(N)$

4 Simulation

Consider single input single output discrete system as follows

$$y(k) = a_1y(k-1)+a_2y(k-2)+b_1u(k-1)+b_2u(k-2)+w(k), k = 0, 1, 2, \cdots, N-1$$

Where, $w(k) \sim N(0, 0.04)$,the parameter$[a_1, a_2, b_1, b_2]$ is the unknown and exists in the model set:$\theta \in \Omega = \{\theta_1, \theta_2, \theta_3\}$, assuming $\theta_1 = [0.8, 0.2, 0.5, 1], \theta_2 =$

$[0.4, 0.6, 0.9, 1], \theta_3 = [0.05, 0.5, 0.2, 1]$. These three models have the same priories probability and all equal to $\frac{1}{3}$. suppose the real parameter of the system is θ_1.Put the above-mentioned equation into the formation of the state equation (5) and (6). It is the aim that seeking control sequence $\{u(k)\}(k = 0, 1, \cdots, N-1)$ and minimizing the performance index expression (4). Fig.1 gives the changing case of the posterior probability of three-models; Fig.2 gives the changing case of the optimal control sequence and dual control sequence. Make 100 Monte Carlo simulation with the method of the suboptimal dual control in the note and that of LQG, the performance indexes are: $J_{dual} = 537.0673$,$J_{opt} = 509.4697$.

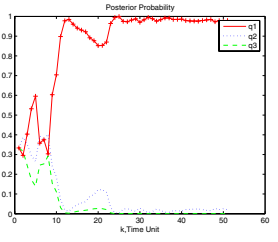


Fig. 1. Posterior probability of the three models

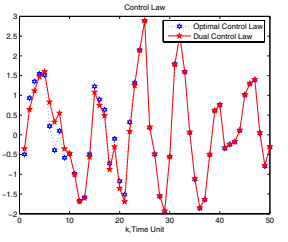


Fig. 2. Comparison of the optimal control sequence and dual control sequence

5 Conclusion

In this paper, we study the dual adaptive control of SISO stochastic system with the unknown parameter by the differential equation. Through transferring the differential equation with the unknown parameter into the state space model, we achieve the suboptimal dual control law. The example is illustrated to clarify the effectiveness and feasibility. This paper is supported by Natural Science Foundation of Shaanxi Province of P.R.China(2005F15).

References

[1] Feldbaum A A. Dual control theory: I-IV[J]. Automat.*Remote Control*, 1960, 21; 1961, 22.
[2] Deshpande J G,Upadhyay T N, Lainiotis D G. Adaptive Control of Linear Stochastic System [J], *Automatica*, 1973, Vol.9: 107-115.
[3] Milito R, Padilla C S, Padilla R A. An innovations approach to dual control[J]. *IEEE Trans. Automatic Control*, 1982, AC-27(1):132-137
[4] Li Duan ,Qian Fucai,Fu Peilin. Variance minimization approach for a class of dual control problems[J]. *IEEE Trans. on Automatic Control*,2002,47(12):2010-2020.
[5] Fu Peilin, Li Duan, Qian Fucai. Active dual control for linear-quadratic Gaussian system with unknown parameters[C]. Proc of the 15th IFAC World Congress. Barcelona,Spain,2002.
[6] Qian Fucai,Liu Ding,Li Yunxia,Dual control based on two-level algorithm[J],*Control Theory and Apllications*,2004,21(1).

Multi-Agent Immune Clonal Selection Algorithm Based Multicast Routing

Fang Liu, Yuan Liu, Xi Chen, and Jin-shi Wang

School of Computer Science and Engineering, Xidian University, Xi'an, 710071, China
f63liu@163.com

Abstract. The least-cost multicast routing with delay constrained is an NP-Complete problems, to deal with which, a Multi-Agent Immune Clonal Selection Algorithm based multicast routing (MAICSA) is proposed in this paper. MAICSA combines the characteristic of Multi-Agent with the search strategy of Immune Clonal Selection Algorithm. To compare with the conventional Genetic Algorithm (GA), MAICSA overcomes the most serious drawbacks, such as slow convergence rate and "prematurity". The experimental results show that MAICSA has faster astringency and higher precision than traditional GA, MAGA (Multi-Agent multicast routing based on Genetic Algorithm) and MAIA (Multi-Agent multicast routing based on Immune Algorithm).

1 Introduction

The term *multicast* means sending information from one source node to multi nodes (but not all) in the communication network [1]. The goal of multicast routing is to find an optimal multicast tree, which satisfies all constraints. The evaluation of finding an optimal multicast tree is an NP-completed problem on both theory and practice, to deal with which, algorithms such as KMB with no constraint and BSMA [1] with constraint have proposed. However, an ubiquitous problem of these algorithms is the failure in keeping the balance of network cost and Quality of Service (QoS) constrains. As an effective stochastic searching algorithm, Genetic Algorithm (GA)[2] can be used to solve multicast routing problem [3][4][5]. However, GA is a general purposed optimization algorithm, where slow convergence rate and precocity are the most serious drawbacks of it. Hence, other many methods have proposed, such as Immune Algorithm (IA) [6] and Immune Clonal Selection Algorithm (ICSA)[7][8], and so forth.

In the last few years, the research of Multi-Agent system has been noticed. A Multi-Agent system [9] is a set of a number of computable agents with intelligent, goals and capabilities. Many researches have shown the advanced performance of agents in the real world. Zhong,W.C., et al presented the Multi-Agent Genetic Algorithm (MAGA) [10]. Liu Jing, et al presented a multiagent evolutionary algorithm for constraint satisfaction problems [11]. Jiao Li-Cheng, et al presented an organizational coevolutionary algorithm for classification [12]. These algorithms are

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 319 – 327, 2006.

practically useful in the optimization of functions with high dimensionality, constraint satisfaction problems and classification. Because finding an optimal multicast tree is considered as the optimization of functions, hence, a Multi-Agent Immune Clonal Selection Algorithm based multicast routing (MAICSA) is proposed in this paper.

2 Multicast Routing Problem Statements

Communication network can be presented as a weighted graph: $G = (V, E)$. Thereof, V is the set of all switching nodes of the network; E is the set of communicating links (x, y) of any two adjacent nodes x and y. $\forall (x, y) \in E$ exist a pair of positive weighted value (*Delay* (x, y),Cost (x, y)) presenting the time-delay and the cost of (x, y) separately. In this paper, we use *Source* presenting the source node of a multicast session, and *DestSet* presenting the set of all members of multicast group. $Source \bigcup DestSet \subseteq V$, $|DestSet|$ is the size of multicast group. Multicast tree $MT = (V_{MT}, E_{MT})$, $(V_{MT} \subseteq V,\ E_{MT} \subseteq E)$. Multicast routing problem i.e. the Steiner tree problem can be declared as looking for a spanning tree which satisfy the following expression:

$$Cost(MT) = \min\left(\sum_{(x,y)\in E_{MT}} cost(x, y)\right) \tag{1}$$

In the route selecting based on QoS problem, the formula (2) is the mathematical model of the multicast routing problem with Delay-Constraint:

$$\sum_{(Source,v)\in Path_{MT}} Delay(source, v) \le \Delta,\ \forall v \in DestSet \tag{2}$$

where $Path_{MT}$ is the set of all paths from source node to all destinations, and Δ is the delay- constraint.

3 Multi-Agent Immune Clonal Selection Algorithm Based Multicast Routing (MAICSA)

A multi-agent system can implement its global goal by these relationships and every agent inside the system implementing its local goals. An Immune Clonal Selection Algorithm has the ability of global searching, it not only overcomes the prematurity to some extent and avoids falling into local optimums. Through a combination of characteristics of a multi-agent system and immune clonal selection algorithm, we proposed MAICSA.

3.1 The Basic Idea of Immune Clonal Selection Algorithm[7]

An Immune Clonal Selection Algorithm [7][8], based on the Antibody Clonal Selection Theory of immune system, solves the optimization problem by constructing the clonal operator corresponds to certain problem. The state transformation of antibody population is denoted as follows:

$$A(k) \xrightarrow{clone} A'(k) \xrightarrow{mutation} A''(k) \xrightarrow{selection} A(k+1)$$

3.2 The Basic Idea of Multi-Agent Genetic Algorithm (MAGA)[10]

A multi-agent system is composed of many autonomic unit agents. The sensation and behavior of an unit agent can only affect a limited local environment. In different problems, agent has different concrete meanings. In this paper, every agent presents one combination of possible paths in the whole network topology, which is a multicast tree. The logic structure of the Multi-Agent system is an L×L agents grid as illustrated in Fig.1, the topology of the grid is annular. Every agent holds a point and cannot move, the parameter range as illustrated in Fig.2 decides the scale of its neighborhood. Since an unit agent has only local sensations, it can only interact with the agents in the neighborhood. In MAGA [10][11], Neighbor-Competition Operator and Agent Crossover Operator are used to realize the competition and cooperation of agents, while the functions of Agent Mutation Operator and Agent Self-study Operator are improving agents' energy with their knowledge. These four operators are leading the evolution of agents. In this paper, mutation is executed by immune clonal operation.

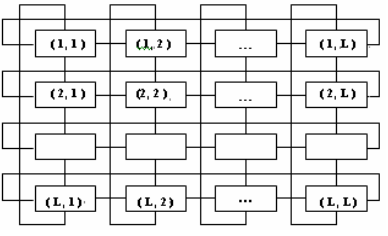


Fig. 1. Logic Structure of Multi-Agent

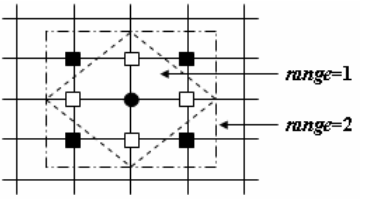


Fig. 2. Neighborhood of the Agent

3.3 Descriptions of MAICSA

Step 1. Generating the set of possible paths

Find Ω_i, the set of possible paths for every pair of notes $(Source, v_i)$ (thereof $v_i \in DestSet$), by Distance Finding Station(DFS). From Ω_i, select the paths satisfying formula (2) and generate Φ_{v_i}.

Step 2. Coding mechanism and initialization

Assume that there are m ($m = |DestSet|$) destinations in the communicational network, each $v_i \in DestSet$ corresponding to the set of possible paths Φ_{v_i}, the number of paths in Φ_{v_i} is n_{v_i}. Code Φ_{v_i} by ordinal positive integers, the proper path code sequence is $\{1, 2, \ldots\ldots, n_{v_i}\}$.

Pick up the number of N_{Pop} (the scale of an antibody population) combinations of paths randomly as an initial antibody community, i.e. *AntibodySet*, within which each antibody presents a combination of paths or a multicast tree. We can code those

antibodies as $a_{v_1} a_{v_2} \cdots a_{v_m}$, the coding length is *m*, and the value of each unit can be:

$$a_{v_i} = \begin{cases} 0, & No\ possible\ path\ of\ \Phi_{v_i}\ be\ selected \\ n_{v_i}, & The\ path\ n_{v_i}\ of\ \Phi_{v_i}\ be\ selected \end{cases} \quad i = 1, 2, \ldots, m \tag{3}$$

Step 3. Calculating the antigen-antibody affinity
Let E_{MT} present the multicast tree corresponding to the $Antibody_i$, the affinity is:

$$f(Antibody_i) = 1/Cost(Antibody_i) = 1\Big/\sum_{(x,y)\in E_{MT}} Cost(x, y) \tag{4}$$

The repetitious link just be calculated once when we calculating $\sum_{(x,y)\in E_{MT}} Cost(x,y)$.

Step 4. Distilling the bacterin
The function of bacterin is, by injecting which, directing search and expediting convergence. In this paper, for each destination, the bacterin is distilled from the shortest path with the lowest cost and common shorter paths in the multicast tree. It's obvious that the method takes advantage of the knowledge of multicast routing problem.
Step 5. Clone Operation
For each antibody $Antibody_i$ of initial population, clone operation duplicates it to the new population $Antibody'_i$ with the scale of q_i, q_i is given by formula (5):

$$q_i = Int\left[N_C \bullet \frac{f(Antigen_i)}{\sum_{j=1}^{N_{Pop}} f(Antigen_j)} \right] \quad i = 1, 2, \ldots\ldots, N_{Pop} \tag{5}$$

where $Nc > N_{Pop}$ is a given integer relating to the clonal scale, $Int\lceil x \rceil = \min\{t > x \mid t \in Z\}$.
As presented in formula (5), the clonal scale of an antibody accommodates to its affinity. The antibody with greater affinity has larger clonal scale.
Step 6. Clonal Mutation Operation
In order to keep the information of the parent population, mutation operation only be applied on the population after Clonal Operation in the Immune Clonal Selection Algorithm. The mutation operation is applied on each unit of every antibody according to the mutation probability P_M.
Step 7. Injecting the bacterin
Pick out some antibodies with lower affinity from the mutated population randomly, they are replaced by the bacterin we distilled in step 4.
Step 8. Clonal Selection Operation
Let $A = \{Antibody_i \cup Antibody'_i\}$, pick out the antibody with the greatest affinity from *A* to express it as B_i, then B_i replaces $Antibody_i$ according to the probability P_T given as formula (6):

$$P_T = \begin{cases} 1 & f(Antibody_i) < f(B_i) \\ \exp(-\dfrac{f(Antibody_i) - f(B_i)}{\alpha}) & f(Antibody_i) \geq f(B_i) \\ & \text{and } Antibody_i \text{ isn't the best of population} \\ 0 & f(Antibody_i) \geq f(B_i) \\ & \text{and } Antibody_i \text{ is the best of population} \end{cases} \quad (6)$$

where $f(\cdot)$ was defined as formula(4) above. $\alpha > 0$, α is related to the diversity of antibody population, in general, the better is the diversity of antibody population, the bigger is value of α.

Step 9. Initializing the Multi-Agent grid

All antibodys are sorted logically. In the Multi-Agent grid, assume the given agent to be $Agent_{ij} = (a_{v_1}, a_{v_2}, ..., a_{v_m})$. Each antibody is put into a certain point of the Multi-Agent grid, that is, $Agent_{11}, \cdots, Agent_{1L} Agent_{21}, \cdots, Agent_{2L}, \cdots, Agent_{L1}, \cdots, Agent_{LL}$ are put into $(1,1), \cdots, (1,L)(2,1), \cdots, (2,L), \cdots, (L,1), \cdots, (L,L)$, as is illustrated in Fig.1.An agent has energy. In the multicast routing problem, let E_{MT} present the multicast tree corresponding to $Agent_{ij}$, $Energy$ ($Agent_{ij}$) is defined as formula (7). Since the goal of MAICSA is to find the lowest cost multicast tree, therefore the goal of the agents' revolution is to increase their selves' energy.

$$Engery\left(Agent_{ij}\right) = 1 \Big/ Cost\left(Agent_{ij}\right) = 1 \Big/ \sum_{(x,y) \in E_{MT}} Cost(x,y) \quad (7)$$

Step 10. Neighbor-Competition Operator

$Max_{ij} = (h_{v_1}, h_{v_2}, ..., h_{v_m})$ is the agent with the highest energy in its neighborhood, namely $Max_{ij} \in Neighbors_{ij}$, and

$$\forall Agent \in Neighbors_{ij},\ Energy(Agent) \leq Energy(Max_{ij}) \quad (8)$$

If formula (9) is true, then $Agent_{ij}$ is the winner of the competition, otherwise it is the loser:

$$Energy(Agent_{ij}) \geq Energy(Max_{ij}) \quad (9)$$

If $Agent_{ij}$ is winner, it will stay alive at its point of the grid. Otherwise it will die and has its point emptied, then Max_{ij} will take over that point using one of the two take-over methods according to the probability P_S.

Let $New_{ij} = (n_{v_1}, n_{v_2}, ..., n_{v_m})$, in the method1, New_{ij} is decided by formula (10):

$$n_{v_k} = \begin{cases} a_{v_k}, & Cost(a_{v_k}) \leq Cost(h_{v_k}) \\ h_{v_k}, & Cost(a_{v_k}) > Cost(h_{v_k}) \end{cases} \quad (10)$$

where $k=1,2,\ldots,m$, and in the method 2, $n_{v_k} = h_{v_k}$, i.e. $New_{ij} = Max_{ij}$.

The two methods have different functions. When $Agent_{ij}$ is loser, it may still have some useful knowledge to the optimization, the method1 benefits keeping that knowledge, the method 2 use the knowledge of the winners to direct searching and discover new search spaces.

Step 11. Agent Crossover Operator

According to the probability P_C, Agent Crossover Operation is applied on $Agent_{ij}$ and Max_{ij} to realize the collaborations among different agents.

Step 12. Agent Self-study Operator

The agents can study with their selves' knowledge to improve their abilities of competing and adapting to the environment. After crossover operator, pick out the agent with the highest energy within the whole agent grid *AgentGrid*, i.e. *Curbest*. $CurBest = (b_1, b_2, ..., b_m)$. According to formula (11), we generates a new Multi-Agent grid *sAgentGrid* with smaller scale (*sL*×*sL*), $sAgent_{i'j'} \in sAgentGrid$, $i', j' = 1, 2, \cdots sL$. The $sAgent_{i'j'}$ is defined by formula (11).

$$sAgent_{i'j'} = \begin{cases} CurBest\ , & i' = 1\ and\ j' = 1 \\ sNew_{i'j'}\ , & otherwise \end{cases} \tag{11}$$

where $sNew_{i'j'}$ is generated by mutation operation on a random single gene unit of *Curbest* according to the probability p'_m . Then the best agent $Curbest'$ will be generated after applying Neighbor-Competition Operator (Step10) and Agent Crossover Operator (Step11) on *sAgentGrid* in presumptive iterative times.

If $Energy(CurBest') > Energy(CurBest)$, then $Curbest'$ replace *Curbest.*

Allbest is the best agent by last iterative time. If *Energy*(*CurBest*)> *Energy*(*AllBest*), then $AllBest \leftarrow CurBest$. MAICSA stops if *Energy*(*AllBest*) doesn't change any more or the iterative times get to the limited number. Else, go back to Step5.

4 Simulations and Results Analysis

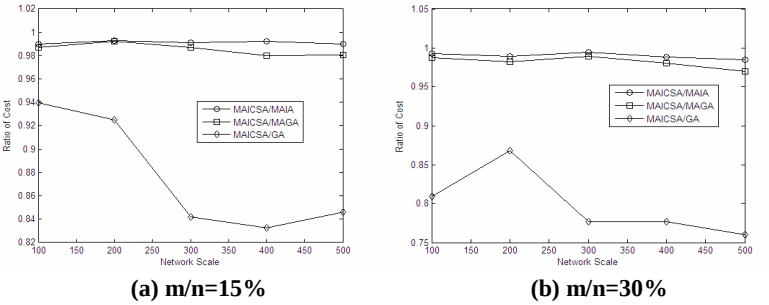
The model of the network we used in our experiment is produced by the method in [13]. In our simulation, we compared the cost of delay-constrained multicast tree got by MAICSA discussed in the paper, MAGA, GA [4] and MAIA [14]. An immune operator is introduced into MAIA, which is to say, distilling the bacterin step and injecting the bacterin step are added to MAGA. To compare easily, we used the same crossover probability P_C =0.8,the mutation probability P_M=0.1 in Clonal Mutation Operation, P_{m1}=P_{m2}=P_{m3}=0.03 in MAGA, MAIA and GA, p'_m = 0.3 and p_s=0.7. The network scale *n* is from 100 to 500, number of destintions is *m* and delay is $\Delta = 120$. The size of *AgentGrid* is 4×4, the size of *sAgentGrid* is 2×2 and rang=1.

The average iterative time of MAIA is bigger than MAGA with the network scale *n* from 100 to 500 because the immune operator (see Table 1), but these differences are very small. The multicast tree cost got by MAIA is lower than MAGA (Cost of MR Tree in Table 1 stands for Cost of Multicast Routing Tree). Because MAISCA is combination of Multi-Agent with immune clonal selection algorithm, it has features of solution diversity and no easy "prematurity". In order to firstly evolve in a larger

population space for the solution diversity, MAICSA has run Clonal Selection Operation before some antibodies selected from the population are put into some points of the Multi-Agent grid. So the average iterative time of MAISCA is the least in GA, MAGA and MAIA with the network scale n from 100 to 500 (as shown Table 1 and Fig.4), the multicast tree cost got by MAICSA is the lowest in GA, MAGA and MAIA with the network scale n from 100 to 500 (see Table 1), Fig.3 (a) and (b) illustrate the multicast tree cost ratio of MAICSA to GA, MAGA and MAIA with m/n=15% and m/n=30%.

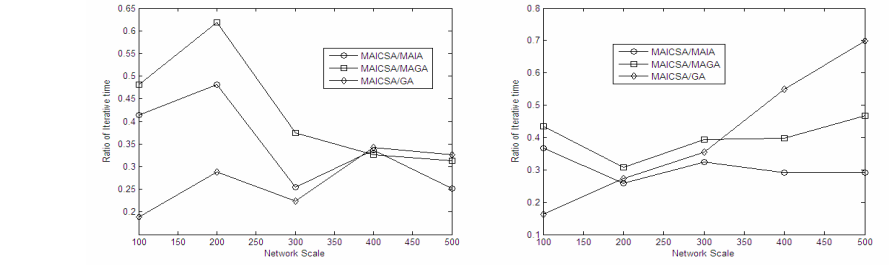
Table 1. The result of four algorithms

Net-work scale	Algor-ithm	*m/n=15%*		*m/n*=30%	
		Cost of MR Tree	Average iterative time	Cost of MR Tree	Average iterative time
100	MAICSA	323	6.9	504.15	9.6
	MAIA	326.25	16.65	507.9	26.15
	MAGA	327.25	14.35	510.5	22.1
	GA	343.85	36.6	623.15	59.1
200	MAICSA	790.65	16.85	1245.4	17.75
	MAIA	796.5	35.05	1259.75	68.45
	MAGA	796.75	27.25	1268.5	57.55
	GA	854.9	58.5	1435	64.85
300	MAICSA	1379.45	18.7	2340.6	28.65
	MAIA	1391.8	73.65	2355.05	88.5
	MAGA	1398	49.85	2366.45	72.75
	GA	1638.35	83.75	3011.85	80.55
400	MAICSA	1598.2	25.3	2733.35	45.6
	MAIA	1611.2	75.15	2766.25	157
	MAGA	1630.5	77.7	2787.8	114.55
	GA	1919	73.95	3520.3	83.05
500	MAICSA	1737.6	29.65	2822.35	66.7
	MAIA	1755.35	118.15	2865.2	229.7
	MAGA	1771.9	94.65	2909.45	142.6
	GA	2053.3	90.95	3714.2	95.5



(a) m/n=15% **(b) m/n=30%**

Fig. 3. The multicast tree cost ratio of MAICSA to GA, MAGA and MAIA



(a) m/n=15% **(b) m/n=30%**

Fig. 4. The average iterative time ratio of MAICSA to GA, MAGA and MAIA

5 Conclusions

In this paper, we have described a new Multi-Agent algorithm based on Immune Clonal Selection Strategy and applied it to the multicast routing problem.

In Multi-Agent system, since there are competitions and cooperations between neighboring agents and the ability of self-studying of agents, we combine these characteristics of Multi-Agent with global powerful searching ability of Immune Clonal Selection Algorithm to solve multicast routing problem. In MAICSA, Agent Mutation Operator is subtituted by Clonal Mutation Operation in step 6.In ICSA (Immune Clonal Selection Alogrithm), Clonal Mutation Operator is the main Operator, so clonal mutation probability is much larger than mutation probability of GA. In consideration of solution stability, the clonal mutation probability of MAICSA is less than one of ICSA, but it is larger than one of GA, MAGA and MAIA .The experiment results have shown that MAICSA has faster astringency and higher precision than traditional GA, MAGA and MAIA. And also, MAICSA can be operated simply and implemented easily. We think that MAICSA is a fruitful approach to study multicast routing problem, it has better prospective of application.

Acknowledgement. This work is supported by the National Natural Science Foundation of China under Grant No. 60372045 and No. 60133010, the Defence Pre-Research Project of China under Grant No.51406020104DZ0124, the Key Science-Technology Project of Higher Education of China under Grant No. 0202A022 and the National Research Foundation for the Doctoral Program of Higher Education of China No. 20030701013.

References

1. Mehrdad Parsa. An interactive algorithm for delay-constrained minimum-cost multicasting. IEEE/ACM Trans on Networking. Vol.6, No.4, (1998) 461-474
2. Chen Guo-Liang e.t. Genetic Algorithm and Its Application. Beijing: People's Posts and Telecommunications Press, (1996)

3. Liu Ying. Multicast routing algorithms computer networks [PhD dissertation]. Xidian University, Xi'an, (2000)
4. Wang Xin-Hong, Wang Guang-Xing. A multicast routing approach with delay constrained minimum cost based on genetic algorithm. Journal of China Institute of Communications, 23(3),pp:112-117,(2002)
5. Shi Jian, Zou ling, Dong Tian-Lin,Zhao Er-Dun. The Application of Genetic Algorithm in Multicast Routing. Acta Electronic Sinica Vol.28 No.5, pp: 88-89 ,(2000)
6. Jiao Licheng and Wang Lei, A novel genetic algorithm based on immune, IEEE Trans. on System, Man, and Cybernetics—Part A, vol.30, pp: 1-10,2000
7. Du Hai-Feng, Jiao Li-Cheng et. Immune clonal selection algorithm and evolutionary algorithms. In 2003 10th International Conference on Artificial Intelligence, Progress of Artificial Intelligence in China (2003) 694-699
8. Jiao Li-cheng, Du Hai-feng. The prospect of the artificial immune system. ACTA ELECTRONICA SINICA.31(10):1540-1548,(2003)
9. Singh M P. Multi-agent system: A theoretical framework for intention, know-how and communication. Berlin:Springer-Verlag, (1944)
10. Zhong Wei-Cai, Liu Jing, Xue Ming-Zhi, Jiao Li-Cheng. A Multi-Agent Genetic Algorithm for Global Numerical Optimization. IEEE Trans.System,Man and Cybernetics—Part B. 34(2),pp:1128-1141 (2004)
11. Liu Jing, Zhong Wei-Cai, Jiao Li-Cheng. A multiagent evolutionary algorithm for constraint satisfaction problems. IEEE Trans. Syst., Man, and Cybern. B, 36(1), pp: 54-73(2006)
12. Jiao Li-Cheng, Liu Jing, Zhong Wei-Cai. An organizational coevolutionary algorithm for classification.IEEE Trans. Evol. Comput. 10(1), pp:67-80(2006)
13. B.M.Waxman. Routing of multiple connections. IEEE Journal of Selected Areas in Communications, Vol. 6, No.9,pp:1617-1622 (1988)
14. Liu Yuan .The Multi-Agent Multicast Routing Algorithm based on Immune Clonal Computation. [MS dissertation]. Xidian University, Xi'an, (2005)

Estimation Distribution of Algorithm for Fuzzy Clustering Gene Expression Data

Feng Liu[1], Juan Liu[1], Jing Feng[1], and Huaibei Zhou[2]

[1] Computer School of Wuhan University, Wuhan University
Wuhan, China
wolflf@126.com, liujuan@whu.edu.cn,
genefeng@mail.whu.edu.cn
[2] International School of Software, Wuhan University, Wuhan, China
bzhou@whu.edu.cn

Abstract. With the rapid development of genome projects, clustering of gene expression data is a crucial step in analyzing gene function and relationship of conditions. In this paper, we put forward an estimation of distribution algorithm for fuzzy clustering gene expression data, which combines estimation of distribution algorithms and fuzzy logic. Comparing with sGA, our method can avoid many parameters and can converge quickly. Tests on real data show that EDA converges ten times as fast as sGA does in clustering gene expression data. For clustering accuracy, EDA can get a more reasonable result than sGA does in the worst situations although both methods can get the best results in the best situations.

1 Introduction

With the rapid advancement of genome sequencing projects, microarrays and related high-throughput technologies have been key factors in the study of global aspects of biological systems. Generally speaking, gene expression data can be gotten from microarray experiments by readout of the mRNA levels of genes, which are conducted to address specific biological questions. The microarray experiments are usually carried on a genome with a number of different samples (conditions) such as different time points, different cells or different environmental conditions [1]. These data are always stored in a matrix, in which each row corresponds to a gene, each column corresponds to a condition, and each entry is a real number and denotes the expression level of the gene under the specific condition. The matrix can be denotes by X(G,C), where G is the set of genes and C is the set of samples. When gene expression data are analyzed, common pursued objectives are to group genes over all samples or cluster samples over all genes. The promise of these objectives is that the similar genes exhibit similar behaviors over all samples, or *vice versa*. This process is called clustering.

A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters [2]. In machine

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 328 – 335, 2006.

learning, clustering is an example of unsupervised learning. Unlike classification, clustering and unsupervised learning do not relay on predefined classes and class-labeled training examples.

There exist a large number of traditional clustering algorithms based on methods of statistics such as partitioning methods, hierarchical methods, density-based methods, grid-based methods and model-based methods. At the same time, many natural computation and other intelligence algorithms, such as neural network, evolutionary computation, fuzzy logic, are introduced into cluster analysis. For an example, fuzzy c means (FCM) is a popular method of cluster analysis and is applied in many fields now.

Combining genetic algorithms and fuzzy logic, zhao *et al* presented a genetic algorithm for fuzzy clustering [3]. However, the behavior of evolutionary computation algorithms such as GAs depends on a large extent on associated parameters like operators and probabilities of crossover and mutation, size of the population, rate of generational reproduction, the number of generations, and so on. So Larranaga and Lozano[4] presented a new algorithm—Estimation of distribution algorithm (EDA)—to come over the disadvantages of genetic algorithm based on probabilistic graphical models. Based on these ideas, we adopt estimation of distribution algorithms for fuzzy clustering gene expression data, combining EDA and fuzzy logic. Experiments on real data show that our methods can outperform sGA both in convergence speed and accuracy of clustering samples.

The remainder of this paper is organized as follows: Model of fuzzy cluster is presented in section 2 and estimation of distribution algorithm is presented in section 3. We do some experiments on real data and discuss the results in section 4. In section 5, we make a conclusion.

2 Model of Fuzzy Cluster

Suppose $X=\{x_1,x_2,\ldots,x_n\}$ is the set of objects to be clustered and each object $x_i=\{x_{i1},x_{i2},\ldots,x_{is}\}$ in X has *s* attributes. The objective of clustering is to partition n objects in X into *c* groups $X_1,X_2,\ldots,X_c$, which satisfy: $X_1 \cup X_2 \cup \ldots \cup X_c=X$, $X_i \cap X_j=\Phi$ for each $1\le i\ne j\le c$.

Traditionally, each object only belongs to one group and each group contains at least one objects. This partition method is called hard partition or crisp partition. Set μ_{ik} to be the grade of membership that object k belongs to group i, then, for crisp partition, the membership function can be denoted as follows:

$$\mu_{ik}=\begin{cases}1 & x_k \in X_i \\ 0 & x_k \notin X_i\end{cases} \tag{1}$$

Ruspini E.[5] introduced the fuzzy logic into partition and define the membershipμ_{ik} to be a real number in [0,1] and satisfy that

$$E_f=\{\mu_{ik} \mid \mu_{ik}\in[0,1];\ \sum_{i=1}^{c}\mu_{ik}=1,\forall k;0<\sum_{k=1}^{n}\mu_{ik}<n,\forall i\} \tag{2}$$

Suppose $U=[\mu_{ik}]_{c\times n}$ is the partition matrix and $P=\{ p_i \mid p_i = (p_{i1},p_{i2},\ldots,p_{is})$ is clustering center (also called clustering prototype) of group i, i=1,2,…, c}. According to Dunn [12] and Bezdek [13], the objective function of clustering is to minimize the following equation:

$$J(U,P)=\sum_{k=1}^{n}\sum_{i=1}^{c}(\mu_{ik})^m(d_{ik})^2 \text{ , s.t. } \mu_{ik}\in E_f \tag{3}$$

where d_{ik} is the distance between the k^{th} object and the i^{th} clustering center and m is weight coefficients. Here m is in $[1,\infty)$ and usually m=2.

According to Zimmermann, H.J[6], if clustering center p_i is known, then fuzzy partition matrix U can be computed as follows:

$$\mu_{ik}=\begin{cases}\dfrac{1}{\sum_{j=1}^{c}(\frac{d_{ik}}{d_{jk}})^{\frac{2}{m-1}}} & if\ I_k=\phi \\ 0,\ \forall i\in \overline{I_k} \text{ and } \sum_{i\in I_k}\mu_{ik}=1 & if\ I_k\neq\phi\end{cases} \tag{4}$$

Where $I_k=\{i|1\le i\le c, d_{ik}=0\}$ and $\overline{I}_k=\{1, 2, \ldots, c\}-I_k$.

In a word, fuzzy clustering is to find a fuzzy partition matrix U or a vector of clustering prototype (clustering center), which can minimize equation (3). In this paper, we try to find the best clustering prototype using estimation of distribution algorithms.

3 Estimation of Distribution Algorithm

3.1 Description of Algorithms

Genetic Algorithm is widely used in searching problems as an optimization algorithm. However, the researcher requires experiences in order to choose the suitable values for the parameters in the algorithm. Therefore, a new type of algorithms, Estimation of Distribution Algorithms (EDAs), was introduced [4]. The algorithms try to make easier to predict the movements of the populations in the search space as well as to avoid the need for so many parameters. Like GAs, the algorithms are also based on the populations and have a theoretical foundation on probability theory.

Unlike GAs, the new individuals in the next population are generated without crossover or mutation operators in EDAs. In fact, they are randomly reproduced by a probability distribution estimated from the selected individuals in the previous generation. At the same time, in EDAs the interrelations between the different variables representing the individuals are expressed clearly by means of the joint probability distribution associated with the selected individuals at each generation.

Suppose a population consists of *R* chromosomes (individuals) and each chromosome is composed of *n* genes as $X_1,X_2,\ldots,X_n$. Then a generic scheme of EDA approaches are shown as figure 1 and the essential steps are following:

Step 1. Generate randomly R individuals, composed of n-dimension, in D_0 generation.
Step 2. Select m ($m<R$) individuals, denoted by D_{h-1}^{m}, from the population in h-1 generation following a criterion.
Step 3. Induce the n-dimensional probabilistic model that better represents the interdependences between the n variables. The model can be presented by a directed acyclic graph (DAG). This is the most crucial step in EDA.
Step 4. Propagate R new individuals, which constitute the new population D_h, by carrying out the simulation of the probability distribution.

Steps 2, 3 and 4 are looped until the terminated condition is satisfied.

There are many ways to estimate the joint probability distribution associated with the selected individuals from the previous generation in discrete domains and more details are introduced in [4].

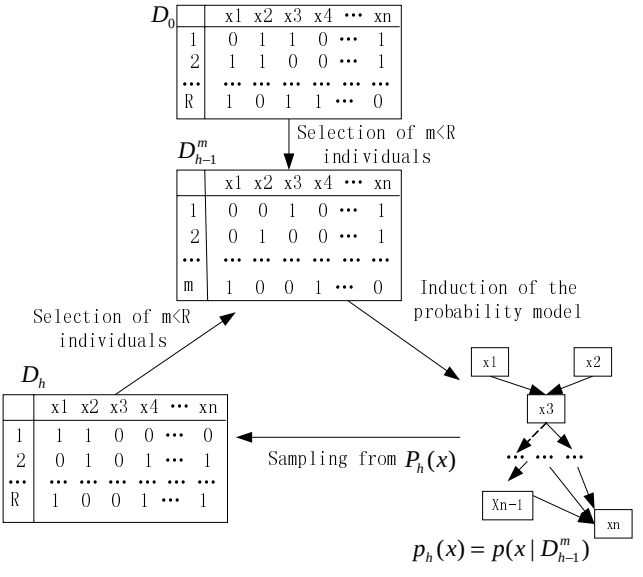


Fig. 1. Illustration of EDA approaches in the optimization process

3.2 Parameter Selection

Each individual in the population denotes clustering prototypes of each partition. Suppose that each object has s attributes and all objects are classified into c classes, then the chromosome is composed of $s*c$ real numbers, which ranges are between the minimum and maximum value of the attribute in all objects. Each real number is discretized into 20-length gray code. Thus each chromosome is coded into $s*c*20$-length gray code. The population size is set to 100 in this paper.

The crucial step in EDA is to induct the probability model from the last populations. In this paper, we suppose that the n-dimensional joint probability distribution factorizes like a product of n univariate and independent probability distribution. So

the famous algorithm UMDA (univariate marginal distribution algorithm) in [4] is adopted to form the directed acyclic graph (DAG) and propagate the next generation in our program.

We use the probabilistic logic sampling (PLS) proposed in [7] to sample the population in next generation. Stochastic Universal Sampling (SUS) is used to select the m individuals from the population in the previous generation. The generation gap in our method is set to 0.95 and the maximum generation is set to 100.

Equation (3) is computed for each individual and taken as the fitness function because our objective is to minimize the equation. In equation (3), d_{ik} is the Euclid distance between k^{th} object and i^{th} clustering center and the fuzzy partition matrix can be computed equation (4).

4 Experiments and Results

We tested our method using two gene expression data sets: ALL/AML data set and human renal tumor data set, which can be downloaded from Gene Expression Omnibus (GEO).

First, we filtered the genes with more than 20 percents null values over all samples. Then, just like [8], we only selected 50 genes that were most informative about the class distinction in the data for each data set. The genes were scored by the "twoing rules" using the tool package Rankgene, which is developed by [9].

The program is developed in MATLAB 7.01 for windows XP using GATBX toolbox, which is developed by University of Sheffield (1994).

To compare with sGA, we also did experiments on the same data sets using sGA. The probabilities of crossover operator and mutation operator were set to be default n GATBX toolbox and other parameters were set as same as UMDA's.

4.1 ALL/AML Data Set

The ALL/AML data set, which is published in [10], consists of the expression levels of roughly 6,800 human genes and data are measured using an Affymetrix oligonucleotide array from bone marrow samples collected from 47 patients suffering from acute lymphoblastic leukaemia (ALL) and 25 patients suffering from acute myeloid leukaemia (AML).

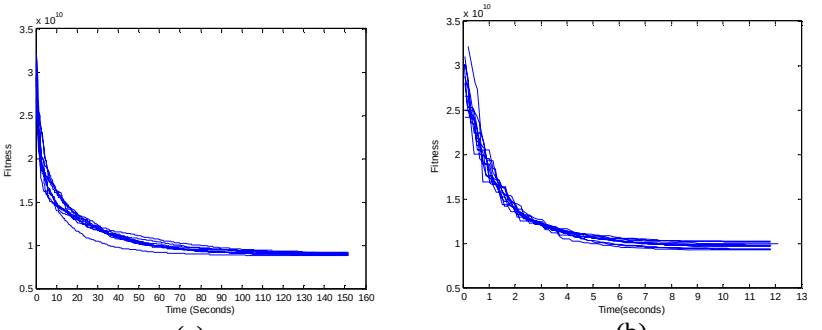
According to the types of patients, we grouped the samples into two clusters using different methods: EDA and sGA, and each run 10 times. The cluster details in the best situations and the worst situations are shown in table 1 and table 2, respectively.

Table 1. Best result for ALL/AML data

	Cluster No.	Cluster samples	AML	ALL	Cluster Error	Total Error
EDA	Cluster 1	24	24	0	0	1/72
	Cluster 2	48	1	47	1/48	
sGA	Cluster 1	24	24	0	0	1/72
	Cluster 2	48	1	47	1/48	

Table 2. Worst result for ALL/AML data

	Cluster No.	cluster samples	AML	ALL	Cluster Error	Total Error
EDA	Cluster 1	26	24	2	2/26	3/72
	Cluster 2	46	1	45	1/46	
sGA	Cluster 1	17	12	5	5/17	18/72
	Cluster 2	55	42	13	13/55	



(a) (b)

Fig. 2. Convergence speed for ten tests on ALL/AML data set. (a) GA's convergence curve (b) EDA's convergence curve.

From table 1, we can we can see that both methods reach the same error rate in the best situation, only one sample is grouped wrongly. The situation emerges 6 times in EDA and 4 times in sGA. In the worst situation listed in table 2 only 3 of 72 samples are wrongly grouped in EDA, while there are 18 of 72 wrongly grouped samples in sGA. That is to say, EDA are quite stable through all runs while the performance of sGA fluctuates across different runs. To sum up, EDA outperforms sGA in clustering ALL/AML gene expression data. Figure 2 shows the converge speeds of two methods. From figure 2, we can see EDA converges in around 5 seconds, while sGA converges in 60 seconds or so. Table 1,2 and figure 2 illustrate that, compared with sGA, EDA can converge more rapidly and get a more reasonable result in practice.

4.2 Kidney Cancer Data Set

Cutcliffe C. *et al* present a human renal tumor data set in 2005, which is composed of 22,283 genes and 35 samples [11]. Excluding three control samples, the 32 left samples come from two types of renal tumor patients: 18 of Wilms' tumor (WT) patients and 14 of clear cell sarcoma of the kidney (CCSK) patients. The expression data are measured using Affymetrix oligonucleotide arrays.

Performing the same experiments as in ALL/AML data set, we found that both methods can group the samples without errors in all runs, with result shown in table 3.

Table 3. Result for Kidney data in all situations

	Cluster No.	cluster samples	WT	CCSK	Error ratio	Total ratio
EDA	Cluster 1	14	0	14	0	0
	Cluster 2	18	18	0	0	
sGA	Cluster 1	14	0	14	0	0
	Cluster 2	18	18	0	0	

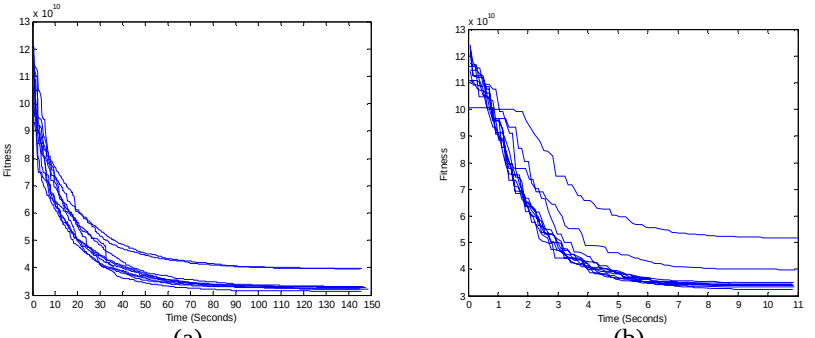


Fig. 3. Convergence speed for ten tests on kidney cancer data set. (a) GA's convergence curve (b) EDA's convergence curve.

The convergence speeds of both methods are shown in figure 3. From figure 3(a), we can find sGA converge slowly and usually converge in 80 seconds or so. However, EDA converge more quickly and can always converge in less than 8 seconds as Fig 3(b) shows. So we can conclude that, for kidney cancer data set, EDA outperforms sGA in convergence speed although both methods can get good result in clustering the samples using gene expression data.

5 Conclusions

In this paper, we put forwarded a fuzzy estimation of distribution algorithm to cluster gene expression data, which combines the estimation of distribution algorithm with fuzzy logic. Our method uses fewer parameters to reproduce the offspring than other evolutionary algorithms which can avoid adjusting the parameters in dealing with different problems. In order to evaluate our method, we clustered two real gene expression data sets using our methods and sGA. For ALL/AML data set, our method outperforms sGA both in accuracy and in convergence speed. For kidney cancer data set, both methods can get the perfect result, but EDA converges much more quickly than sGA. That is to say, in both real data, EDA outperforms sGA in convergence speed and gets a more reasonable result than sGA does. In this paper, we just assume that the variables are independent and use UMDA model to compute the joint

probabilities in EDA, in the future, we will consider the interrelations between variables and introduce more sophisticated joint probability models. Much more comparison experiments with other EAs are also needed to evaluate the overall performance of our method.

Acknowledgement

This paper is supported by the National Nature Science Foundation of China (60301009), Chenguang Project of Wuhan city (211121009).

References

1. Baldi, P., Hatfield, G.W.:DNA Microarrays and Gene Expression: From Experiments to Data Analysis and Modeling, Cambridge Univ. Press.(2002)
2. Han, J,. kamber, M.: Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers (2000):355-395.
3. Zhao, L. , Tsujimura, Y., Gen, M.: Genetic Algorithm For Fuzzy Clustering, International Conference on Evolutionary Computation (1996): 716-719
4. Larranga P., Lozano, J.A.: Estimation of Distribution Algorithms: a New Tool For Evolution Computation. Kluwer Academic Press, Boston, (2001)
5. Ruspini, E.: A New Approach to Clustering, Inf. Control, vol. 15 (1969):22--32
6. Zimmermann, H.J.: Fuzzy set Theory and Its Applications, 4thed. Kluwer Academic Publishers, (2001).
7. Henrion,. M: Propagation of Uncertainty In Bayesian Networks By Probabilistic Logic Sampling. In Uncertainty in Artificial Intelligence 2, Elsevier, North-Holland (1988): 149--163.
8. Murali, T. M., Kasif,S: Extracting Conserved Gene Expression Motifs From Gene Expression Data, In PSB, vol 8 (2003)
9. Su Y., Murali T.M., Pavlovic,V. *et al*: RankGene: Identification of Diagnostic Genes Based on Expression Data, Bioinformatics Vol. 19 No. 12 (2003): 1578-1579
10. Golub,T.R.: Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. Science, Vol 286. No. 15 (1999):531-537
11. Cutcliffe, C., Kersey, D., *et al:* Clear Cell Sarcoma of The Kidney: Up-regulation of NeuralMarkers With Activation of The Sonic Hedgehog and Akt Pathways. Clin Cancer Res vol 11 No.22 (2005):7986-7994
12. Dunn,J.C., A graph theoretic analysis of pattern classification via Tamura's fuzzy relation. IEEE Trans. SMC, Vol 4,No.3 (1974): 310-313.
13. Bezdek,J.C., Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York, (1981).

A Maximum Weighted Path Approach to Multiple Alignments for DNA Sequences

Hongwei Huo[1], Vojislav Stojkovic[2], and Zhiwei Xiao[1]

[1] School of Computer Science and Technology, Xidian University
Xi'an 710071, China
hwhuo@mail.xidian.edu.cn
[2] Department of Computer Science, MSU, Baltimore
Maryland 21251, USA
stojkovi@jewel.morgan.edu

Abstract. This paper presents a novel approach, called MWPAlign-Maximum Weighted Path approach to multiple ALIGNment, to perform global multiple alignment of DNA sequences. In our method, de Bruijn graph is used to describe input sequences information. As a result, a consensus-finding problem can be transformed to a maximum weighted path problem of the graph. MWPAlign gets almost linear computation speed of multiple sequences alignment problem. Experimental results show that the proposed algorithm is feasible, and for large number of sequences with lower mutation rate 5.2%, MWPAlign generates better alignment and has a lower computation time as compared to CLUSTALW, T-Coffee and HMMT.

1 Introduction

The Multiple Sequence Alignment (MSA) is one of the challenging tasks in bioinformatics. It is computationally difficult and has diverse applications in sequence assembly, sequence annotation, structural and proteins, phylogency and evolutionary analysis.

The best-known system based on progressive multiple alignment is perhaps CLUSTALW[1]. Other multiple alignment systems that are mostly targeting proteins or short DNA sequences, and based on progressive alignment, include T-COFFEE[2], MAFFT[3], MUSCLE[4], and PROBCONS[5]. DIALIGN[6]uses iterative refinement strategy for multiple alignments. The partial order alignment (POA) approach[7] models a multiple global alignment as a directed acyclic graph that includes blocks similarity and regions of non-homology. The A-Brujin graph (ABA) approach further generalizes POA to allow an alignment to be represented by a graph that can include cycles, allowing for the inclusion if repeated structure[8].

In this paper, A new method graphed based was presented for multiple sequence alignment, called MWPAlign, in which de Bruijn graph was applied to global multiple alignment for DNA sequences, so that the time and space complexity is almost reduced to linearity, $O(NL)$. MWPAlign is tested for simulated sequences. Experimental results show that MWPAlign is feasible, and for

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 336–339, 2006.

large number of sequences with mutation rate less than 5.2%, MWPAlign generates better alignment results and has lower computation speed compared with CLUSTALW, T-Coffee and HMMT.

2 Multiple Sequence Alignments

Let s_1, s_2, ... ,s_n be n sequences of length l_1, l_2, ..., l_n over an alphabet , which is $\{A, C, G, T\}$, for DNA sequences and twenty different amino acid for protein sequences. Let Σ' denote the extended alphabet $\Sigma \cup \{-\}$, where $'-'$ denote the gap character (reserved for insertions and deletions). A multiple alignment of sequences s_1, s_2, ... ,s_n is specified by a $n \times m$ matrix S, where $max_{1 \leq i \leq n} l_i \leq m$. Each element of the matrix is a member of S', and each row i contains the characters of s_i in order, interspersed with $M - l_i$ gaps. We also assume that every column of the multiple alignment matrix contains at least one symbol from Σ , that is, no column in a multiple alignment contains only gaps. The multiple alignment matrix we have constructed is a generalization of the pairwise alignment matrix to $n > 2$ sequences. The score of a multiple alignment is the weighted sum of scores of all the induced pairwise alignments, with the optimal alignment being the one that maximizes the score.

3 Algorithms

The main idea of the algorithm MWPALign is to produce consensus sequence from the maximum weighted path, and then compare each of the sequences with the consensus sequence using pairwise alignment, finally construct the global multiple alignment of the sequences unambiguously.

3.1 Algorithm Design

The algorithm MWPAlign for multiple sequence alignment can be formally described as follows.

Algorithm MWPAlign
Input $S = s_1, s_2, ..., s_n$, each s_i has length l_i
Output $S' = s'_1, s'_2, ..., s'_n$, each s_{ri} has length m
1. use S to construct de Bruijn graph $G = (V, E)$
2. eliminate cycles of G
3. get a maximum weighted path from G, then construct consensus sequence s_c from the path
4. **FOR** $i \leftarrow 1$ **to** n
5. **DO** $s'_i \leftarrow pairwisealignment(s_c, s_i)$
6. construct output result $S' \leftarrow s'_1, s'_2, ..., s'_n$

The first step is to construct a directed de Bruijn graph using S. This might leads to a vertex with more than one incoming or outgoing edges and may result in cycles in the graph when the number of incoming or outgoing edges in a vertex

increases. So the second step is to remove cycles in the graph and transform it into a directed acyclic graph while keeping all the similarity information among sequences as much as possible. The third step is to compute a path with maximum weighted value on the directed graph, then extract the consensus path from the maximum weighted path. The for loop in lines 4 and 5 repeatedly performs banded pairwise alignment between the consensus path and each input sequence. The last step is to form the final multiple alignment in terms of the pairwise alignment.

3.2 Algorithm Analysis

Let N denote the number of sequences to be aligned, L denote the average sequence length and k denote the size of the k-structure used in construction of the graph.

MWPAlign runs approximately in linear time $O(NL)$, where NL is the total size of the sequences. The steps of the algorithm MWPALign each runs in almost linear time. During the construction of the graph, notice that the total number of the k-structure is $N * (L - k + 1)$, so this step takes the running time of $O(NL)$ in the worst case. We use DFS to transform a graph, while the running time of DFS is the $O(V' + E')$, where V' and E' is the set of vertices and edges in the transformed graph, respectively. There are at most $N * (L - k + 2)$ vertices and $N * (L - k + 1)$ edges in the transformed graph. Thus, this step has the running time $O(NL)$. During the process for solving maximum weighted path, both the topological sort and shortest-path finding in DAG are performed in time $O(V' + E')$, that is, $O(NL)$. In the last step, we use banded dynamic programming to perform local alignment and yield a banded matrix for sequence alignment. This strategy can cut off part of the matrix and leave only a diagonal band, which can reduce the time complexity of the sequence alignment algorithm to be linear, also the $O(NL)$.

Hence, the total running time of MWPAlign algorithm is approximately in linear time $O(NL)$, where NL is the total size of the sequences.

4 Experimental Results

MWPAlign has been implemented on Pentium 4 1.5GHz under the linux operating system with C language for simulated sequences. We generate the tasted data in the following way: first yielding a sequence at random, and then performing mutation with a probability of 5.2%, 10.5%, and 16.4% on the sequence, respectively.

Table 1 shows the results from MWPAlign, CLUSTALW, T-COFFEE, and HMMT with mutation rate 5.2%, with various N, L. We use "F" to represent the fail alignment because T-COFFEE doesn't works for the number of sequences being tested more than 60. The digits in parenthesis denote time(by second) used by that approach.

Table 1. Multiple alignment comparison among MWPAlign, CLUSTALW, T-Coffee, and HMMT with mutation rate 5.2%

N	L	MWPAlign	CLUSTALW	T-Coffee	HMMT
10	100	0.857/0.664(1)	0.844/0.581(1)	0.830/0.542(2)	0.515/0.103(1)
10	200	0.831/0.630(2)	0.823/0.561(2)	0.806/0.519(4)	0.510/0.078(1)
10	500	0.846/0.657(2)	0.833/0.584(4)	0.816/0.543(9)	0.530/0.095(2)
50	100	0.746/0.191(3)	0.728/0.171(4)	0.703/0.093(100)	0.522/0.008(2)
50	200	0.741/0.158(7)	0.732/0.148(9)	0.706/0.064(223)	0.730/0.013(3)
50	500	0.740/0.193(17)	0.700/0.149(55)	0.671/0.065(623)	0.461/0.010(7)
100	100	0.688/0.100(6)	0.655/0.098(10)	F/F(F)	0.474/0.014(2)
100	200	0.714/0.107(13)	0.666/0.105(35)	F/F(F)	0.490/0.004(5)
100	500	0.700/0.101(36)	0.641/0.094(185)	F/F(F)	0.401/0(20)
500	500	0.640/0.090(227)	0.561/0.091(4656)	F/F(F)	0.273/0(90)

5 Conclusions

A new algorithm graph-based called MWPAlign is presented for aligning multiple DNA sequences. We found the maximum weighted path that corresponds to the consensus sequences of the input sequences without doing pairwise alignments. It yields good results with a lower computational time.

References

1. Julie DThompsonDesmond GHigginsToby JGibson(1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22),4673-4680
2. Notredame, C., Higgins, D. G. and Heringa, J.(2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, 302,205- 217
3. Katoh, K., Misasa, K., Kuma, K. and Miyata, T.(2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.*, 30(14), 3059-3066
4. Edgar, R. C(2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, 32(5), 1792-1797
5. Do, C. B., Brudno, M. and Batzoglou, S.(2005). ProbCons: Probabilistic consistency-based multiple alignment of amino acid sequences. *Genome Research*, 15,330-340
6. Burkhard Morgenstern(2004). DIALIGN: multiple DNA and protein sequence alignment at BibiServ. *Nucleic Acids Res.*, 32,W33-W36
7. Lee C, Grasso C and Sharlow M.F(2002). Multiple sequence alignment using partial order graphs. *Bioinformatics*, 18(3),452-464.
8. Raphael B, Zhi D, Tang H and Pevzner P(2004). A novel method for multiple alignment of sequences with repeated and shuffled elements. *Genome Research*,14(11): 2336-2346

Accelerating the Radiotherapy Planning with a Hybrid Method of Genetic Algorithm and Ant Colony System*

Yongjie Li and Dezhong Yao

School of Life Science and Technology,
University of Electronic Science and Technology of China,
610054 Chengdu, China
{Liyj, Dyao}@uestc.edu.cn

Abstract. Computer-aided radiotherapy planning within a clinically acceptable time has the potential to improve the therapeutic ratio by providing the optimized and customized treatment plans for the tumor patients. In this paper, a hybrid method is proposed to accelerate the beam angle optimization (BAO) in the intensity modulated radiotherapy (IMRT) planning. In this hybrid method, the genetic algorithm (GA) is used to find the rough distribution of the solution, i.e., to give the initial pheromone distribution for the following ant colony system (ACS) optimization. Then, the ACS optimization is implemented to find the precise solution of the BAO problem. The comparisons of the optimization on a clinical nasopharynx case with GA, ACS and the hybrid method show that the proposed algorithm can obviously improve the computation efficiency.

1 Introduction

The goal of radiotherapy is to deliver a required dose to the tumor, while sparing those neighboring organ-at-risks (OARs) and normal tissues as much as possible. Intensity-modulated radiotherapy (IMRT) is a powerful mean to potentially improve the therapeutic ratio by using modulated beams from multiple directions to irradiate the tumors. The conventional IMRT planning starts with the selection of suitable beam angles, followed by an optimization of beam intensity maps under the guidance of an objective function and the prescribed dose distributions [1][2].

Beam angle optimization (BAO) is an important but also challenging issue in IMRT planning because of the inherent complexity, mainly the large search space and the coupling between the beam configuration and the beam intensity maps [3][4]. Many published papers have shown that BAO is most valuable for a plan with a small number of beams (<5) [1], and is also clinically meaningful for plans with large number of beams (>9) in complicated cases, where the tumor volume surrounds a critical organ, or is surrounded by multiple critical organs [3][5].

At present, beam angle selection is generally based on the experience of the human planners in the clinical practice. Several trial-and-error attempts are normally needed in order to find a group of acceptable beams [3]. To date, though there are fruitful

* This work is supported by a grant from the 973 Project of China (Grant No. 2003CB716106) and grants from NSFC of China (Grant No. 30500140 and 30525030).

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 340 – 349, 2006.

improvements achieved [3~11], computer-aided BAO still cannot act as a routine planning tool because of the limitation of the extensive computation time.

Mathematically, any appropriate objective function as a function of beam orientations is nonconvex, which requires the use of either an exhaustive search or stochastic optimization methods in order to deal with the potential multiple extrema [1][5]. Among those published algorithms introduced in radiotherapy planning, there exist a number of natural algorithms, which borrow the intelligence from the nature organisms and inherit a global search mechanism. Two typical examples are simulated annealing (SA) [3] [10] and genetic algorithm (GA) [4][7][11].

Since 1996, a new population-based algorithm named ant colony optimization (ACO) has gradually gained much favor [12]. First proposed by Colorni, Dorigo and their colleagues (1991), ACO algorithms are a class of heuristic search algorithms biologically inspired from the behavior of colonies of real ants, and in particular how they forage for food [12~14]. To date, ACO has demonstrated significant successes in dealing with the difficult combinatorial optimization problems in many fields including, among others, traveling salesman problem (TSP) [15], graph coloring problem [16], routing for telecommunication networks [17], quadratic assignment problem (QAP) [18], data mining [19], etc.

In our previous works we have preliminary shown the feasibility of ACO for the BAO problem [20~21], and the limited results showed the better convergence of ACO over GA. The reasons could be partially explained as follows. The ACO algorithm is based on a simplified model of the swarm theory. In the socio-cognitive viewpoint, this means that the mind and, as a consequence, the intelligence, are social attributes. ACO is distinctly different from GA in that it constructs an entire new solution set (colony) in each generation, while GA focus on improving the set of solutions or a single solution from previous iterations. The constructive cooperation between all ants in a generation provides more chances for ants in the swarm to share preferable information between them compared to the GA, in which the information sharing only happens between two parent individuals in a generation.

It has been normally accepted that, GA suffers from slow convergence before providing an accurate solution, i.e., GA has a poor ability of exploitation [22~23]. However, GA turns out to be very good for exploration [24~25]. On the contrary, ACO has a good ability of exploitation, but the canonical ACO suffers from the relatively lower convergence because the pheromone distribution is simply initialized randomly, i.e., there is little initial information on the path. In this paper, a new hybrid method, named GA-ACO, is presented to accelerate the convergence of BAO by combining the merits of GA and ACO. The basic idea is to use GA to generate preliminary BAO results, converting them into initial pheromone distribution for ACO, and then using ACO to search for optimal beam angles.

The rest of this paper is organized as follows. After a brief description of the BAO problem and the general ideas of the proposed algorithm, the mathematical implementation of BAO with the hybrid algorithm is presented in details. A clinical nasopharynx tumor case is studied in order to investigate the efficiency of the proposed algorithm by comparing with the separate GA and ACO algorithms. Finally, some conclusions and directions for future research are presented in the last section.

2 Materials and Methods

2.1 Problem Descriptions and General Ideas

In radiotherapy with external beams, a specified number of beams are used to irradiate the tumor from different directions. IMRT uses intensity-modulated beams to produce expected doses by dividing these beams into beamlets (also called rays) and calculating the weights of these beamlets with different optimization algorithms.

In essence, BAO is a combinational optimization problem, in which a specified number of beams are to be selected among a beam candidate pool. As for the coplanar radiotherapy, the beam directions cover the 360º gantry rotational angles while the collimator angle and couch angle keep fixed. In BAO implementation the continuous 360º angles are generally discretized into equally spaced directions with a given angle step, such as 5º. As for each trial beam combination, the beam intensity maps are optimized to get the corresponding dose distribution inside the body volume, and this dose distribution is used to evaluate the performances of the trial solution.

Based on our previous works [7][20~21], here we propose a hybrid algorithm with GA and ACO to accelerating the solution of BAO. In this hybrid algorithm, GA is used to find a group of rough solutions (chromosomes) for BAO, and then these sub-optimized solutions are converted to the initial pheromone values of the ACO model, according to the fitness values of each chromosome. Then the solution is fine-tuned by ACO. The simplified flowchart is graphically illustrated in Fig. 1.

Due to the limited space, here we do not restate the detailed implementations of BAO with GA [7] and BAO with ACO [21~22]. The details could be found in our corresponding works. In the following two sections we just summarize these two algorithms for the convenience of understanding the proposed hybrid algorithm.

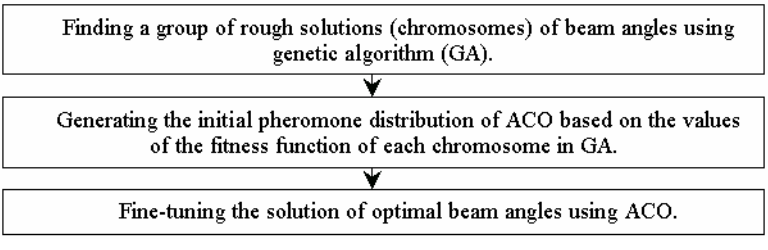


Fig. 1. The simplified flowchart of the proposed hybrid algorithm

2.2 Finding the Rough Solution with Genetic Algorithm

The GA adopts a one-dimensional integer-coding scheme, in which the beam combination is represented as a chromosome (an individual), and the length of chromosome is equal to the beam number of the plan [7]. Each gene represents a trial angle. The individual *parent 1* shown in Fig. 2 demonstrates a five-beam plan with angles: 0°, 80°, 140°, 190° and 300°. The genes in one chromosome are different with each other, which means that there should be no two beams with same angles in one plan.

Four genetic operations are adopted in this study, i.e., selection, crossover, mutation, and pseudo immunity (Fig. 2). Parent individuals with higher fitness are selected into the next generation with a higher probability. To the selected parent individuals, a crossover operation is applied according to a crossover probability. Then a mutation operation to the children will be done according to a mutation probability. Finally a pseudo immunity operation is applied to the children sets. For example, in the first individual generated after the mutation operation, the angles in the first and last genes are 310° and 300°, respectively, which are two angles with too small separation for an acceptable plan. So it is reasonable to replace one of them with a suitable angle (310° is randomly selected to be randomly replaced by 220° in Fig. 2).

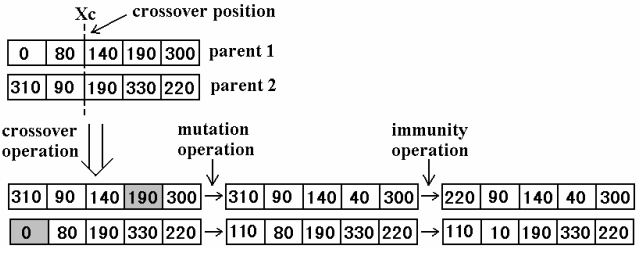


Fig. 2. The coding scheme and the genetic operations for beam angle optimization

2.3 Fine-Tuning the Solution with ACO

The first ACO algorithm, called Ant System (AS), was initially introduced to solve TSP [14]. Subsequently, some modifications were carried out over the original AS. Among those ant colony meta-heuristics, the ant colony system (ACS) algorithm is a typical improvement of the AS algorithm.

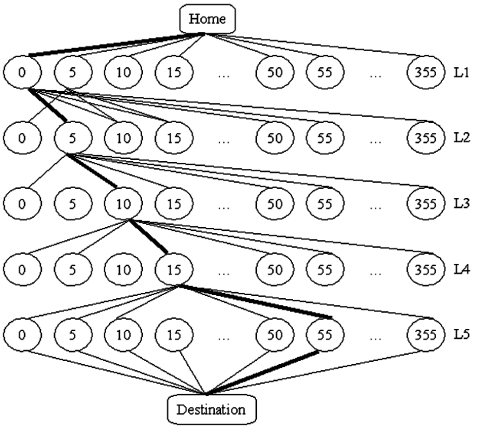


Fig. 3. The graph representation of the BAO problem with the ACO algorithm [20]

Fig. 3 shows the graph representation of the BAO problem using ACS [20]. In this graph, the layer number is equal to the beam number, and the node number at each layer is equal to the number of the angle candidates. Each node is associated with a discrete angle. Considering a five-beam plan shown in Fig. 3, the layer number is 5. The total 360º angles are discretized into equi-spaced directions with a step of 5º, and then the node number at each layer is 72. The thick black lines show a solution constructed by an ant, which includes angles of 0º, 5º, 10º, 15º, and 55º.

Instead of randomly initializing the pheromones on the edges in Fig. 3, the proposed algorithm initializes them according to rough solutions given by each individual in GA, and the amount of the pheromone on the edge described by this individual is proportional to the fitness value of the individual. The pheromones of the remaining edges not being covered by those individuals are initialized randomly. After the initialization of pheromone, the best solution is to be search by ants cycle by cycle. The general process of the ACS-based optimization could be described as follows.

A colony consists of *N* ants. The ants start from the *home* node and move from the first layer to the last layer in each cycle. Each ant can select only one node at each layer according to the state transition rule given by Eqs. (1) and (2), which give the probability with which ant k in node i chooses to move to the node j on the next layer. The best state is chosen with probability q_0 (that is a parameter $q_0 \in [0,1]$ usually set to 0.8~0.9) according to

$$p_k(i,j)=\begin{cases}1, & \text{if } j=\arg\max\limits_{r\in A_k(i)}\left\{[\tau_{ir}]\cdot[\eta_{ir}]^{\beta}\right\}\\ 0, & \text{if } j\neq\arg\max\limits_{r\in A_k(i)}\left\{[\tau_{ir}]\cdot[\eta_{ir}]^{\beta}\right\}\end{cases} \tag{1}$$

and with probability $(1-q_0)$ the next state is chosen randomly with a probability distribution based on τ_{ij} and η_{ij} according to

$$p_k(i,j)=\begin{cases}\dfrac{[\tau_{ij}]\cdot[\eta_{ij}]^{\beta}}{\sum\limits_{r\in A_k(i)}[\tau_{ir}]\cdot[\eta_{ir}]^{\beta}}, & \text{if } j\in A_k(i)\\ 0, & \text{otherwise}\end{cases} \tag{2}$$

Where τ_{ij} is the pheromone intensity on edge (i,j); η_{ij} is the heuristic function value of the edge (i,j); $A_k(i)$ is the set of angles that remain to be visited by ant k positioned on node i, and $\beta>0$ is a parameter which determines the relative importance of pheromone versus heuristic function. Here, the heuristic function value η_{ij} is calculated based on the beam's-eye–view dosimetrics (BEVD) score given by [10]

$$\eta_{ij}=\frac{1}{N_T}\sum_{n\in tumor}\left(\frac{d_{nj}}{D_T^P}\right)^2 \tag{3}$$

Where d_{nj} is the maximum dose delivered to the voxel n by the *j*th beam; N_T is the number of voxels in the tumor, and D_T^P is the prescription dose to the tumor.

While constructing a feasible solution (i.e., a trial plan), ants visit edges and change their pheromone level by applying the local updating rule given by

$$\tau_{ij} \leftarrow (1-\psi)\cdot\tau_{ij} + \psi\cdot\tau_0 \tag{4}$$

Where $0<\psi<1$ is a parameter and τ_0 is the initial pheromone level.

Once all ants have completed their tours, a *cycle* is finished and a global pheromone updating rule is applied, and only the globally best ant is allowed to deposit pheromone. The global updating rule is given by

$$\tau_{ij} \leftarrow (1-\rho)\cdot\tau_{ij} + \rho\cdot\Delta\tau_{ij} \tag{5}$$

$$\Delta\tau_{ij} = \begin{cases} \dfrac{\zeta\cdot F_{\max}}{F_{obj}}, & \text{if } (i,j)\in \text{ global best tour} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

Where $0<\rho<1$ is a pheromone decay parameter. F_{obj} is the objective function value, and $F_{\max}$ is a estimation of the maximum value of the objective function, which is used to keep the global updating more controllable by ensuring the value of $F_{\max}/F_{obj}$ around 1.0. ζ is a parameter to control the scale of the global updating of pheromone.

2.4 Objective Function and Fitness Value

For each trial plan, a conjugate gradient (CG) method is used to optimize the corresponding beam intensity maps [2][7], and then the dose distributions calculated using these optimized intensity maps are used to evaluate the individual. The optimization aims to minimize the dose difference between the prescribed and calculated dose distributions, which can be simply described by the following objective function

$$F_{obj}(\vec{b}) = \sum_{j=1}^{NT} \delta\cdot w_j\cdot\left(d_j(\vec{b}) - p_j\right)^2 \tag{7}$$

Where $F_{obj}(\vec{b})$ is the value of objective function of the beam set $\vec{b}$, NT is the point number in the volume. $\delta = 1$ when point dose in the volume breaks the user-specified constraints, else $\delta = 0$; w_j is the weight of the *j*th point, d_j and p_j are the calculated and prescribed doses of the *j*th point in the volume, respectively.

As for the BAO problem with GA, the quality of each trial plan is evaluated by a fitness value, which is calculated by

$$Fitness(\vec{b}) = F_{\max} - F_{obj}(\vec{b}) \tag{8}$$

Where $F_{\max}$ is a rough estimation of the maximum value of the objective function. Both $F_{\max}$ and $F_{obj}(\vec{b})$ are calculated using Eqs. (7) ~ (8).

3 Results

In order to verify the feasibility of the proposed algorithm (GA-ACS, for short), three types of optimization tasks are separately implemented with GA, ACS and GA-ACS, based on a clinical case with nasopharynx tumor (planning tumor volume, PTV) shown in Fig. 4(a). There are three organ-at-risks (OARs) to be protected during the irradiation: spinal cord, the left and right parotids. Five 6MV coplanar photon beams are used to irradiate the tumor.

Experimentally and statistically, the number of ants is set to 30, the coefficient β in Eqs. (1) and (2) is set to unity, q_0 is fixed to 0.8, the parameter ψ in Eq. (4) is set to 0.6, and the pheromone decay parameter ρ in Eq. (5) is set to 0.6. A population size of 30 is used for GA, same as the number of ants in ACO. 0.95 and 0.01 are used for the crossover and mutation probability of GA, respectively. The angle search space is discretized with angle step of 5º, which results in 72 discrete angles in the candidate pool among the whole 360º coplanar gantry angle.

For a relatively fair comparison, all of the three methods run independently 20 times. All of the runs find the same optimal beam angles: 10°, 80°, 155°, 205° and 280°, shown as the dotted straight lines in Fig. 4(b). The result of BAO is compared with that of a manual plan with a set of five equi-spaced beams with the angles of 0º, 72º, 144º, 216º and 288º. The beam directions and the doses of the manual plan are shown in Fig. 4(a). From these dose distributions we can see that, with a higher conformality, the PTV in the optimal plan is more tightly surrounded by high doses, and the OARs are spared high doses for larger volumes, especially for the spinal cord and the right parotid, compared to the dose distributions of the manual plan.

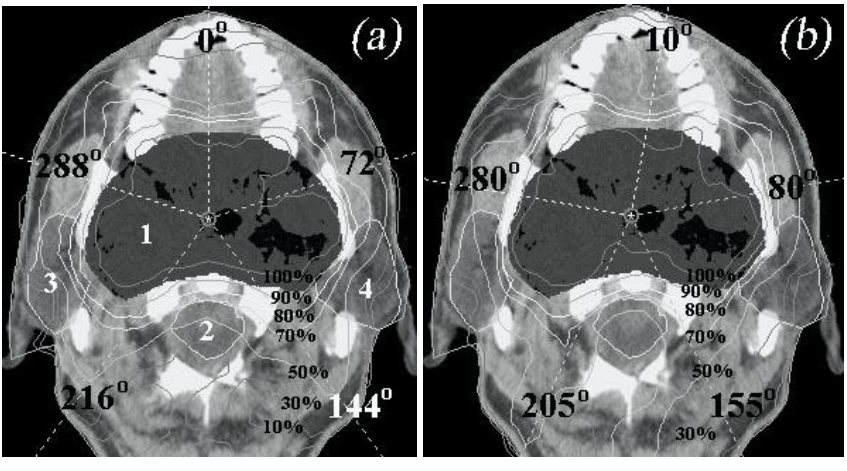


Fig. 4. The beam angles and dose distributions of (a) the manual plan and (b) the optimal plan. 1-tumor (PTV), 2-spinal cord, 3-right parotid, 4-left parotid.

The comparisons about the computation efficiency are listed in table 1. Compared to the GA-based optimization algorithm, the ACS algorithm reduces the mean computation time from 28 min 43 s to 23 min 16 s, about 19% of the computation time is saved; the hybrid GA-ACS algorithm further reduces the mean computation time to 18 min 22 s, about 21% of the mean computation time is meaningfully saved.

The improvement of computation efficiency of GA-ACS is also graphically illustrated by the curves of fitness value versus generation number shown in Fig. 5, in which each of the three curves is the one with minimum computation time among 20 runs for the three algorithms, and the fitness values are those of the current best individual. The faster convergence of GA over ACS in the earlier stage and the quicker convergence of ACS over GA in the later stage of the optimization progress are obviously illustrated in the figure. In the GA-ACS algorithm, the first 20 generations are implemented by GA, and the later optimization progress after the 20th generation is conducted by ACS. The improvement in computation efficiency of GA-ACS over GA and ACS is also clearly shown in Fig. 5, i.e., GA-ACS has a better convergence than that of GA and ACS.

Table 1. The statistical comparisons about the computation efficiency of the three algorithms

Optimization algorithm	Run times	Minimum optimization time	Maximum optimization time	Mean optimization time
GA	20	26 min 57 s	30 min 27 s	28 min 43 s
ACS	20	22 min 33 s	26 min 08 s	23 min 16 s
GA-ACS	20	16 min 32 s	20 min 19 s	18 min 22 s

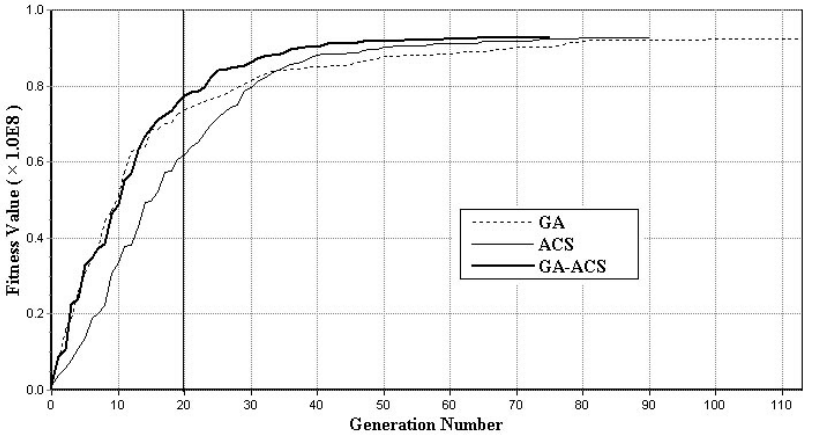


Fig. 5. The fitness value versus generation number curves of the three algorithms. The curves are from the runs with the minimum computation time of each algorithm. Note that the fitness values are those of the current best individual (plan).

4 Discussion and Conclusions

In this paper, a new optimization technique (named GA-ACS) was introduced to accelerate the beam angle optimization (BAO) in IMRT planning by hybridizing the GA and ACS algorithms. The basic idea of GA-ACS is to use GA to generate rough BAO results, which are converted into initial pheromone distribution for ACS, and then ACS is followed to fine-tune the optimal beam angles. This hybrid method combines the merits of the faster convergence of GA in the earlier progress and the better convergence of ACS in the later progress of the optimization. The efficiency comparison on a clinical nasopharynx tumor case shows that all of the GA, ACS, GA-ACS can produce a better dose distribution by optimizing the beam angles compared to the manual plan. More importantly, the GA-ACS-based BAO algorithm improves the optimization efficiency meaningfully compared to the separate GA and ACS.

Beam angle optimization for IMRT planning is a difficult problem because of the extensive computation. Many efforts are still expected before the automatic selection of beam angles becomes a routine tool. The limited results in our works reflect an interesting potential of ACO meta-heuristic approach for the BAO problem and encourage further research on the topic.

However, it should be noted that, though there are so many published works on the efficiency comparisons between GA and ACO, we still couldn't conclude that one algorithm is absolutely preferable to another in any field at any circumstance. Each algorithm holds its own merits. It is now an obvious trend to hybridize different algorithms by combining their merits together.

The results in this paper are preliminary, and two directions for future research are as follows. First, more test cases are needed to thoroughly compare the performance of GA-ACS, which may help us improve its performance. Second, it is necessary to make an insight into the heuristic function. Knowledge-based solution is an inevitable trend for those computer-assisted diagnosis and therapy systems. The probabilistic state transition rule in ACS provides us a chance to incorporate the human experience and knowledge about the specified engineering problem into the optimization process by formulating them into a heuristic function.

References

1. Webb S.: Intensity-modulated Radiation Therapy. Bristol and Philadelphia, Institute of Physics Publishing (2000)
2. Spirou S. V., Chui C. S.: A gradient inverse planning algorithm with dose-volume constraints. Med. Phys. 25 (1998) 321-333
3. Pugachev A., Boyer A. L., Xing L.: Beam orientation optimization in intensity-modulated radiation treatment planning. Med. Phys. 27 (2000) 1238-1245
4. Hou Q., Wang J., Chen Y., Galvin J. M.: Beam orientation optimization for IMRT by a hybrid method of genetic algorithm and the simulated dynamics. Med. Phys. 30 (2003) 2360-2376
5. Gaede S., Wong E., Rasmussen H.: An algorithm for systematic selection of beam directions for IMRT. Med. Phys. 31 (2004) 376–388
6. Djajaputra D., Wu Q., Wu Y., Mohan R.: Algorithm and performance of a clinical IMRT beam-angle optimization system. Phy. Med. Biol. 48 (2003) 3191–3212

7. Li Y.J., Yao J., Yao D.Z.: Automatic beam angle selection in IMRT planning using genetic algorithm. Phy. Med. Biol. 49 (2004) 1915–1932.
8. Souza W. D., Meyer R. R., Shi L.: Selection of beam orientations in intensity-modulated radiation therapy using single-beam indices and integer programming. Phy. Med. Biol. 49 (2004) 3465–3481
9. Wang X., Zhang X., Dong L., Liu H., Wu Q., Mohan R.: Development of methods for beam angle optimization for IMRT using an accelerated exhaustive search strategy. Int. J. Radiat. Oncol. Boil. Phys. 60 (2004) 1325–1337
10. Pugachev A., Li J. M. S., Boyer A.L., et al.: Role of beam orientation optimization in intensity-modulated radiation therapy. Int. J. Radiat. Oncol. Boil. Phys. 50 (2001) 551–560
11. Schreibmann E., Xing L.: Feasibility study of beam orientation class-solutions for prostate IMRT. Med. Phys. 31 (2004) 2863-2870
12. Dorigo M., Stützle T.: The ant colony optimization metaheuristic: Algorithms, applications and advances. In Handbook of Metaheuristics, volume 57 of International Series in Operations Research & Management Science, F. Glover and G. Kochenberger Eds, Norwell, MA: Kluwer Academic Publishers, 2002, pp. 251-285
13. Colomi A., Dorigo M., Maniezzo V.: Distributed optimization by ant colonies. in Proc. First Europ. Conf Artificial Life, Varela F and Bourgine P, Eds. Paris, France: Elsevier, 1991, pp. 134-142.
14. Dorigo M., Colorni A., Maniezzo V.: The ant system: Optimization by a colony of cooperating agents. IEEE Trans. Syst. Man Cybern. B, 26 (1996) 29-41.
15. Dorigo M., Gambardella L. M.: Ant colony system: a cooperative learning approach to the traveling salesman problem. IEEE Transactions on Evolutionary Computation, 1 (1997) 53-66
16. Costa D., Hertz A.: Ants can color graphs. Journal of the Operational Research Society, 48 (1997) 295-305
17. Di Caro G., Dorigo M.: AntNet: Distributed stimergetic control for communication networks. Journal of Artificial Intelligence Research, 9 (1998) 317-365
18. Maniezzo V and Colorni A: The ant system applied to the quadratic assignment problem. IEEE Trans. Knowledge and Data Engineering, 11 (1999) 769-778
19. Parpinelli R. S., Lopes H. S., Freitas A. A.: Data mining with an ant colony optimization algorithm. IEEE Transactions on Evolutionary Computing, 6 (2002) 321-332
20. Li Y.J., Yao D.Z., Chen W.F., Zheng J.C., Yao J.: Ant colony system for the beam angle optimization problem in radiotherapy planning: A preliminary study. 2005 IEEE Congress on Evolutionary Computation Proceedings (CEC'2005), 2 (2005) 1532-1538
21. Li Y.J., Yao D.Z., Yao J.: Optimization of Beam Angles in IMRT Using Ant Colony Optimization Algorithm. International Journal of Radiation Oncology, Biology, Physics, 63 (2005) Supplement 1 (ASTRO'2005) S492-S493
22. Renders J. M., Flasse S. P., Hybrid Methods Using Genetic Algorithms for Global Optimization. IEEE Transactions on Systems, Man, and Cybernetics-part B: Cybernetics, 26 (1996) 243-258
23. Christopher M. C., Edward J. R., John E. R., Investigation of Simulated Annealing, Ant-Colony Optimization, and Genetic Algorithms for Self-Structuring Antennas. IEEE Transactions on Antennas and Propagation, 52 (2004) 1007-1014
24. Yao X.: Evolutionary Computation: Theory and Applications. World Scientific, Singapore (1999)
25. Tan K.C., Lim M. H., Yao X., Wang L. P. (Eds.): Recent Advances in Simulated Evolution And Learning. World Scientific, Singapore (2004)

Model Deconstruction of an Immunoprevention Vaccine*

F. Pappalardo[1,2], P.-L. Lollini[3], S. Motta[2], and E. Mastriani[2]

[1] Faculty of Pharmacy, University of Catania
[2] Department of Mathematics and Computer Science, University of Catania
francesco@dmi.unict.it, motta@dmi.unict.it, mastriani@dmi.unict.it
[3] Sezione di Cancerologia, Dipartimento di Patologia Sperimentale,
and Centro Interdipartimentale di Ricerche sul Cancro "Giorgio Prodi"
University of Bologna
pierluigi.lollini@unibo.it

Abstract. We present a further step in developing an *in silico* model that simulates the immune system responses to tumor cells in vaccinated mice. Our study is based on our previous model that simulates the cancer - immune System competition activated by a tumor vaccine. In what follows, we show first results on effects of deconstructing the components of Triplex vaccine that are in very good agreement with *in vivo* results.

1 Introduction

Immunoprevention of mammary carcinoma in HER-2/neu transgenic mice was attempted using various immunological strategies, including cytokines, non-specific stimulators of the immune response, and HER-2/neu specific vaccines made of DNA, proteins, peptides, or whole cells.

A new vaccine, called Triplex, was proposed in [3]. The vaccine combined three different stimuli for the immune system. The first was p185neu, protein product of HER-2/neu, which in this system is at same time the oncogene driving carcinogenesis and the target antigen [1].

A complete prevention of mammary carcinogenesis with the Triplex vaccine was obtained when vaccination cycles started at 6 weeks of age and continued for the entire duration of the experiment, at least one year (Chronic vaccination) [2].

We developed an accurate model [6] of immune system responses to vaccination. We performed *in silico* experiments considering a large population of individual mice. Each individual mouse is characterized by a sequence of uniformly

* This work was supported in part by IMMUNOGRID project, under EC contract FP6-2004-IST-4, No. 028069. F.P. and S.M. acknowledge partial support from University of Catania research grant and MIUR (PRIN 2004: *Problemi matematici delle teorie cinetiche*). This work has been done while F.P. is research fellow of the Faculty of Pharmacy of University of Catania. P.-L.L. acknowledges financial support from the University of Bologna, the Department of Experimental Pathology ("Pallotti" fund) and MIUR.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 350–353, 2006.

distributed random numbers which will determine the probabilistic events. As showed in [6], comparison with *in vivo* experiments shows excellent agreement. The model has been already used in searching for better therapies [4,7].

In this paper we present a further, yet important, step in developing the model of immune system response to vaccination. In particular we show initial results on the effect of the three components of the Triplex vaccine. *In silico* surviving results are in good agreement in *in vivo* experiments [5]. This result enhances the model and its potential applications.

2 Materials and Methods

We have used our previous model [6] as a starting point. We have enhanced the SimTriplex simulator in order to have more accurate results. Refinements included: *i)* a better and precise implementation of allogeneic MHC factor in the simulator. *ii)* Interleukin 12 management; it has been introduced separately in the interactions dendritic cells $\leftrightarrow$ helper T cells and in vaccine cells $\leftrightarrow$ helper T cells; here interleukin 12 allows a vaccine cell to have the same behavior of an antigen presenting cell, stimulating humoral response; natural killer cells $\leftrightarrow$ cancer cells; here we allow a natural killer cell to kill a cancer cell coated by antibodies only if there is a sufficient amount of interleukin 12, we have set up accordingly to *in vivo* experiments. *iii)* a better and precise implementation of tumor associated antigens (TAA) and specific p185neu.

These enhancements have been allowed us to reproduce some *in vivo* results from [5].

3 Results

In setting the computer experiments, we randomly chose 10 virtual mice over the 100 of the first sample set we used in [6]. All the 10 virtual mice gave very similar, but obviously not identical, results. Figure 1 shows *in vivo* results using a Chronic schedule with different Triplex's components combinations. First two cases has been already simulated in [6].

With refinements introduced in SimTriplex simulator, we were able to successfully reproduce third (a vaccine with only the Tumor Associated Antigen, p185) and fourth case (a vaccine containing only the allogeneic class I major histocompatibility complex (MHCI) glycoproteins) of figure 1.

The third case on figure 1 shows that mice treated with a Triplex vaccine that lacks interleukin 12 and allogeneic class I major histocompatibility complex (MHCI) glycoproteins have the same behavior of the no treated mice. We perfectly reproduce this behavior in our *in silico* experiments. Figure 2 (right side) shows that cytotoxic T cells response is almost flat. This is due to fact that a vaccine without allogeneic MHCI and interleukin 12 components is unable to stimulate both the cytotoxic response driven by cytotoxic T cells and humoral response driven by T helper cells. The Tumor Associated Antigen is a *self* and

p185	Allo-MHC	IL-12	Mammary ca. median latency (wk.)
+	+	+	>52
–	–	–	20
+	–	–	20
–	+	–	20
–	–	+	27
+	+	–	28
–	+	+	25
+	–	+	25

Fig. 1. Triplex vaccine effects. A "+" sign means that the vaccine components is present. A "-" sign means that the vaccine components has been omitted [5].

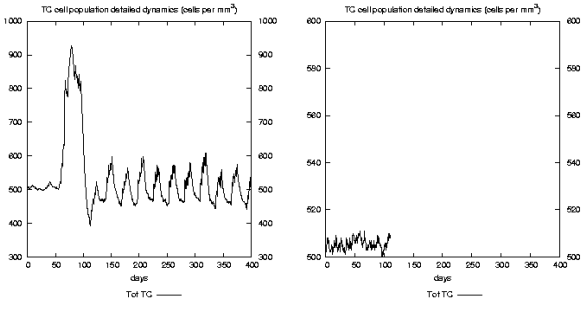


Fig. 2. Cytotoxic T cells behavior in absence of IL12 and allogeneic MHCI vaccine components (left side) and with full Triplex vaccine (right side)

will not be recognized. Figure 2 (left side) shows cytotoxic T cells response with normal Triplex vaccine.

The fourth case reproduces the Triplex vaccine without interleukin 12 and p185neu specific antigen. Our virtual mice produce solid tumor in about 20 weeks, in exact agreement with *in vivo* experiments. As shown in figure 3 (right side), specific antibodies were not produced by immune system. This is due to missed stimulation of helper T cells from B cells. Antigen presenting cells (B cells, macrophages, dendritic cells) were not be able to capture and process

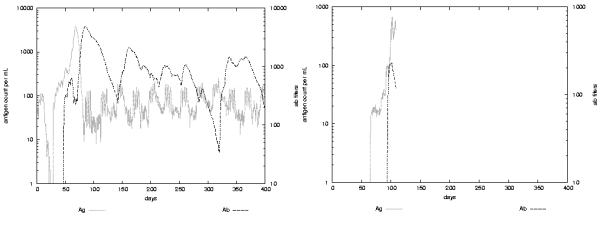


Fig. 3. Antigens - antibodies behavior with Triplex lacking IL 12 and p185neu components (right side) and full components (left side)

tumor associated antigens (p185neu) that are not released by vaccine cells. Helper T cells are not stimulated and consequently B cells don't duplicate and don't become plasma B cells able to release specific antibodies. Moreover the absence of interleukin 12 did not allow direct stimulation of helper T cells from vaccine cells. As one can see, left side of figure 3 show a constant presence of specific antibodies in mice treated with full Triplex vaccine.

4 Conclusion and Future Work

We have presented a further step in developing a model that is able to reproduce *in vivo* experiments on HER-2/neu transgenic mice describing the action of a tumor vaccine in stimulating immune response and the ensuing competition between the immune system and tumor cells.

Introduced refinements demonstrate that the model and its computer implementation are very flexible and new biological entities, behavior, and interactions can be easily added. This helps achieving a realistic description of the immune responses that target solid tumor formation.

We plan to further investigate in improving the model, taking into account new biological requirements.

References

1. Colombo, MP., Trinchieri, G.: Interleukin-12 in anti-tumor immunity and immunotherapy. Cytokine Growth Factor Rev., **13:2** (2002) 155-168
2. De Giovanni, C., Nicoletti, G., Landuzzi, L., Astolfi, A., Croci, S., Comes, A., Ferrini, S., Meazza, R., Iezzi, M., Di Carlo, E., Musiani, P., Cavallo,F., Nanni, P., Lollini, P.-L.: Immunoprevention of HER-2/neu transgenic mammary carcinoma through an interleukin 12-engineered allogeneic cell vaccine. Cancer Res., **64:11** (2004) 4001-4009.
3. Lollini, P.-L., De Giovanni, C., Pannellini, T., Cavallo, F., Forni, G., Nanni, P.: Cancer immunoprevention. Future Oncology, **1:1**, (2005) 57-66
4. Motta, S., Lollini, P.-L., Castiglione, F., Pappalardo, F.: Modelling Vaccination Schedules for a Cancer Immunoprevention Vaccine. Immunome Research, **1:5**, (2005)
5. Nanni, P., Nicoletti, G., De Giovanni, C., Landuzzi, L., Di Carlo, E., Cavallo, F., Pupa, S.M., Rossi, I., Colombo, M.P., Ricci, C., Astolfi, A., Musiani, P., Forni, G., Lollini, P.-L.: Combined Allogeneic Tumor Cell Vaccination and Systemic Interleukin 12 Prevents Mammary Carcinogenesis in HER-2/neu Transgenic Mice. J. Exp. Med. **194**, (2001) 1195-1205
6. Pappalardo, F., Lollini, P.-L., Castiglione, F., Motta, S.: Modelling and simulation of cancer immunoprevention vaccine. Bioinformatics, **21:12**, (2005) 2891-2897
7. Pappalardo, F., Mastriani, E., Lollini, P.-L., Motta, S.: Genetic Algorithm against Cancer. LNCS, **3849**, (2006) 223-228

Detection of Individual Microbubbles Using Wavelet Transform Based on a Theoretical Bubble Oscillation Model

Yujin Zong, Bin Li, Mingxi Wan*, and Supin Wang

Department of Biomedical engineering, Key Laboratory of Biomedical Information Engineering of Ministry of Education, Xi'an Jiaotong University, 710049, Xi'an, China
{yjzong, ornot, mxwan, spwang}@mail.xjtu.edu.cn

Abstract. Detecting individual microbubbles is important for the quantification of the amount of bubbles in the tissues, determination of microvascular volume and targeted microbubble imaging. We took the advantage of a theoretical bubble oscillation model to construct a matched wavelet, *i.e.* bubble wavelet as mother wavelet to detect individual microbubble using wavelet transform. The experimental echoes with different levels of added noises were processed. The results showed significant improvement even for an Echo-Noise-Ratio (ENR_{in}) of -20 dB and the spatial location demonstrated very close agreement with the original experimental echo. This technique was much better than those based on harmonic analysis especially under the circumstance of short pulse insonation.

1 Introduction

Bubble detection is a challenging topic in many industrial, medical and environmental applications[1]. During the past decades, several acoustic detection techniques have been developed based on the strong backscattering properties of bubbles[2][3]. Usually, ultrasound contrast agents (UCAs) are encapsulated microbubbles similar in size to human erythrocytes[4]. Based on the linear, nonlinear or transient characteristics of UCAs, several UCAs specific detection method have been developed to improve the ultrasound images quality[5][6]. These detection techniques are mainly prone to improve the image quality, that is, to detect the group effect demonstrated by microbubble population rather than individual microbubbles. Besides, the very weak echo backscattered from individual microbubble makes it difficult to detect.

However, in recent years, the detection of individual microbubbles is gradually attracting researchers' attentions due to the emerging of "Molecular Imaging". Unger[7] and Lamerichs[8] have reported that nowadays ultrasound imaging is highly sensitive to small amounts of contrast microbubbles – it is even possible to detect an individual microbubble. The capability of ultrasound imaging to detect individual microbubbles is important for the quantification of bubbles in the tissues, determination of microvascular volume and targeted microbubble imaging. Klibanov has reported the

* Corresponding author.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 354 – 363, 2006.

images of individual microbubbles presented as write foci in B-mode image 8, wherein a very low concentration ~90-240/ml was used, that is more less than the usual concentration used in UCAs specific detection modalities, ~10^6/ml or even more. A clinic application was also reported by Dorset[10][11]. They employed contrast transcranial Doppler ultrasound (TCD) to diagnose the right-to-left shunts by detecting the presence and the number of individual microbubbles within intracranial arteries.

A higher spatial resolution is required for detection of individual microbubbles since the detection target is so tiny. Accordingly, the length of insonation burst should be short. However, a very short insonation burst goes against the onset of the harmonic frequencies of individual microbubbles, especially of the subharmonic and superharmonic frequencies[12][13][14]. Even we can augment the amplitude of the harmonic frequencies simply by increasing the amplitude of insonation pressure to an enough high level, the risk of the destruction of microbubbles will make these harmonic techniques based on Fourier transform (FT) fail to function.

Considering the "pulse-like" echo of a microbubble insonified by a single incident burst, detecting individual microbubbles may be equivalently thought as detecting the response of a microbubble from the background noise. If the information about the desired waveform can be known in advance and used in detection, the optimization of the performance of detection will be achieved[15].Several theoretical models on the dynamics of single oscillating microbubble have been well established, which can be used to predict the echo backscattered by a single microbubble in our detection. The basic idea of the method based on matched filter is to calculate the correlation between prior known echoes, so called matched filter, predicted by using the theoretical model and a detected signal, and then find the maximum in the correlation. The occurrence time of the maximum will demonstrate the presence of microbubble. Intuitively, a matched filter can be interpreted as a wavelet-based detector with the predicted echo waveshape of an oscillating microbubble serving as a mother wavelet function. The inherent shift and scale properties of wavelet transform (WT) can provide more flexible and powerful performance than common correlation analysis[16].

2 Method

A simplified model is used to represent the corresponding signals:

$$x(t) = A \cdot s\left[\frac{t-b}{a}\right] + n(t)\,. \tag{1}$$

Where $x(t)$ is the signal received by the transducer used for detection; $A \cdot s\left[\frac{t-b}{a}\right]$ is the desired echo of insonified microbubble, which is time and amplitude scaled and is a time translated version of a predicted waveform, $s(t)$. Wherein A is an amplitude scaling factor, a is a time scaling factor and b is a time delay (or shift); $n(t)$ is the noise component.

2.1 Modified Herring Model

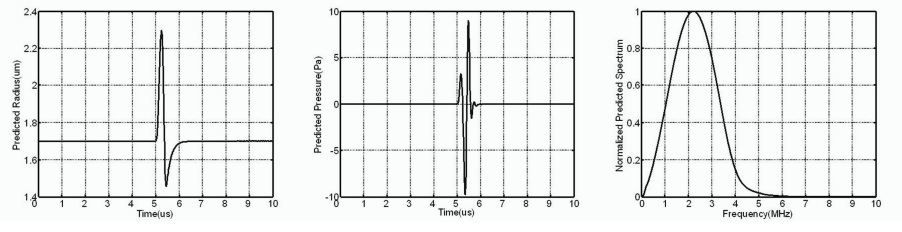
A modified Herring equation was employed to predict the radius-time curve and furthermore the echo backscattered of microbubbles[17][18]:

$$\rho R\ddot{R}+\frac{3}{2}\rho\dot{R}^2=\left(P_0+\frac{2\sigma}{R_0}+\frac{2\chi}{R_0}\right)\left(\frac{R_0}{R}\right)^{3k}\left(1-\frac{3\gamma}{c}\dot{R}\right)-\frac{4\mu\dot{R}}{R}-\frac{2\sigma}{R}\left(1-\frac{1}{c}\dot{R}\right).$$
$$-\frac{2\chi}{R}\left(\frac{R_0}{R}\right)^2\left(1-\frac{3}{c}\dot{R}\right)-12\mu_{sh}\varepsilon\frac{\dot{R}}{R(R-\varepsilon)}-\left(P_0+P_{driv}(t)\right) \quad (2)$$

The predicted echoes were calculated as follow:

$$P(t)=\rho r^{-1}(R^2\ddot{R}+R\dot{R}^2)\,. \quad (3)$$

Where r is the distance to receive the echo from the bubble and $s_{pred}(t) = P(t)$.



(a) (b) (c)

Fig. 1. (a) Theoretical radius-time curve, (b) Predicted echo, (c) Normalized amplitude spectrum. of a insonified microbubble (R_0=1.7 μm, χ = 0.26 *N/m*, ε = 10 *nm*, μ_{sh} = 1.49 *Pa.s*, *P=486kPa*)

Figure 1 shows a typical result predicted by the modified Herring equation, which includes the radium-time curve, the echo pressure waveform, e.g. *s(t)* and the corresponding normalized spectrum.

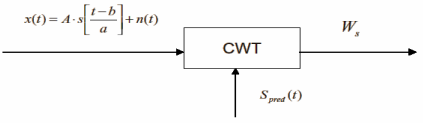


Fig. 2. Schematic representation of the detection method

2.2 Matched Wavelet

A method formulated in terms of match wavelet is used, as depicted in Fig. 2, where $S_{pred}(t)$ represents the prior information as described in Eq. (3).

For one-dimensional data, the continuous wavelet transform (CWT) maps the time variable into a two-dimensional signal with parameters a, scale, and b, shift 19.The continuous wavelets transform is given by

$$W_s(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} s(t) \cdot \psi_{a,b}^*(t) dt \,. \tag{4}$$

Where

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi(\frac{t-b}{a}) \,. \tag{5}$$

Here $\psi(t)$ is an appropriate square integrable function termed as mother wavelet, which is chosen to analyze a specific transient signal of finite energy. The transformed result $W_s(a,b)$ is called wavelet transform coefficients.

Theoretically, the mother wavelet can be any function as long as it satisfies the following admissibility condition:

$$\int_{-\infty}^{+\infty} \frac{|\psi(\omega)|^2}{\omega} d\omega < +\infty \,. \tag{6}$$

where $\psi(\omega)$ is the Fourier transform of the mother wavelet. In our detection practice, according to the frequency response of the experimental system, the predicted echoes are bandpass filtered with 100 KHz-10 MHz, and this results in the consistence with the fact that for Eq. (6) to be satisfied, $\psi(\omega)$ must have a bandpass like spectrum and has a zero direct current (D.C.) value i.e. $\psi(0) = 0$. Even though the admissibility condition can not be well satisfied by the predicted echo, which means the transformation is not energy persevering, the amplitude of the coefficients will be proportional to the energy of the predicted echo, which is our really concern in the detection.

3 Experiment and Data Analysis

3.1 Materials and Protocol

The experimental contrast agent was prepared by sonicating the surfactants mixture solution with SF_6 gas[19]. The mean radius of microbubble was $1.7\mu m$ and 82% microbubbles were in the radius range of $0.5\mu m$ to $3.5\mu m$, the concentration was approximately 3.6×10^8/ml. Before injection, the sample was further diluted with phosphate buffered saline (PBS) solution (pH=7.4).

The experimental acoustical setup was shown in Fig. 3. Arbitrary wave generator (AWG2021, Sony/Tektronix, JP) produced a single sinusoid bursts with the frequency of 2.25MHz at a pulse repeated frequency (PRF) of 1 Hz for transmission. The transmitted sinusoid pulses were first amplified in a linear power amplifier (25A250A, AR, USA) and then supplied to a transmitting transducer (V306; Panametrics,

Waltham, MA). The signals scattered from microbubble were received by a receiving transducer (V309; Panametrics) and amplified with a RF amplifier (BR640, Retic, USA). The amplified signals were acquired using a digital oscilloscope (TDS 340A, Tektronix, USA) and then transferred via a GPIB to a PC for further processing.

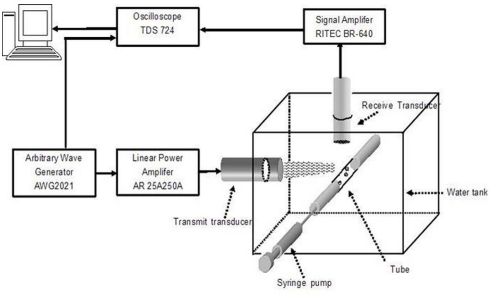


Fig. 3. Acoustical measurement diagram

The acoustic pressures of the co-focal region of the transmitter and receiver were measured with a needle hydrophone (HPM02/1; Precision Acoustics Ltd.). A cellulose tube with an inner diameter of 200 μm was located at the co-focal region, which was connected to a syringe pump (Pump 11, Harvard Apparatus, USA) to control the flow rate of the microbubbles. To ensure only one microbubble in the sample volume at a given time, the suspension was injected at a flow rate of 1 ml/h. A mini magnetic stirrer was used in the syringe to avoid the microbubbles' floating up during the injection.

3.2 Data Analysis

To verify the robustness of the method, various levels of the Gaussian white noise *W(n)* were added to the received signal *E(n)* and their relationship was represented by an input Echo-Noise-Ratio, ENR_{in}:

$$ENR_{in} = 10\log\frac{\sum_{n=1}^{N_1}(E(n)^2/N_1)}{\sum_{n=1}^{N_2}(W(n)^2/N_2)}. \tag{7}$$

The received signals with added noises were analyzed using CWT and then produced two-dimension signals with parameters (*e.g.* shift and scale), representing the time-scale distribution of the signals. The intensity of the two-dimension signals, i.e. the WT coefficients, represented the degree of the correlation between the received signal and the predicted echo, that is, the maximum of the WT coefficients should be the position indicator of microbubble. In practice, however, we adopted the -3 dB value of the maximum as the threshold to discriminate the microbubble from noises and this leaded to a contour-line representation in the two-dimension plane. In some sense, the time-span of the area encircled by the contour lines can be thought as the temporal resolution of the method.

For the convenience of comparision, we also defined ENR_{out} to represent the output Echo-Noise-Ratio. The scale S_{max} at which the maximum of WT coefficients was found is picked to obtain a one-dimension signal with a parameter (*e.g.* shift) and hence obtaining ENR_{out}. Since synthetic signals was used in verifying and approximately knowing the location and the length of the signal, the ENR_{out} can be calculated as the ratio of the output signal power computed in a time window where the received echo was present and a time window for which the original echo was null[21]. It should be pointed out that the "echo" in ENR_{out} was pseudo signal, actually the WT coefficients; additionally, it should be kept in mind that the scale in calculating the ENR_{out} was always $S_{max.}$

3.3 Results

Figure 4 (a)-(d) showed for different ENR_{in}, the typical experimental echo of microbubble with additive Gaussian white noises (the top) and corresponding wavelet transform results: 2D-contour lines of WT coefficient, scale S and shift time (the middle); 1D-coefficient line at the maximum of S, S_{max} (the bottom). The 2D-contour

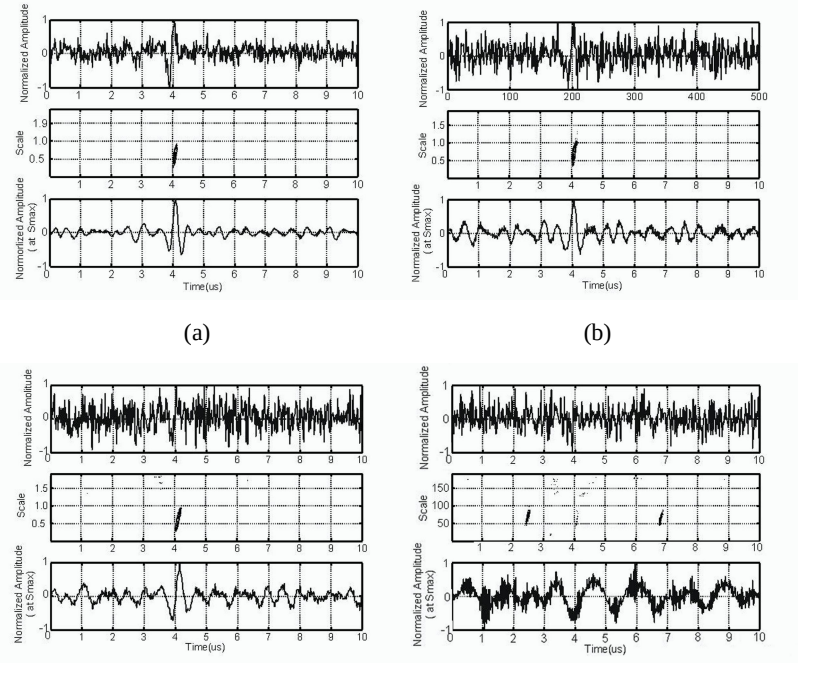


Fig. 4. Typical experimental echo with additive noises (the top), the contour plot of WT coefficients (the middle) and the coefficient line at scale S_{max} (the bottom) for different ENR_{in} (a) -0.86 dB (b)-20.32 dB (c)-23.39 dB (d) -31.86 dB

lines only denoted the -3db value of the maximum of the wavelet coefficients and above. It was clearly shown that with the decreasing of $ENR_{in,}$ the experimental echo of microbubble was submerged in the noises and unable to be visually discriminated any more. However, with the detector based on the bubble wavelet as demonstrated in Fig. 1, the submerged echo was revealed accurately again in the contour depiction and 1D-coefficient line at S_{max} even when ENR_{in} was -23 db. With the further decreasing of ENR_{in}, the method based on the bubble wavelet did not work well any longer and the probability became little that the maximum of the wavelet coefficients fall in the interval where the experimental echo was actually presented, that is, it was hard to assert the location of microbubble echo in statistics sense. Fig. 5 showed a statistics results of 100 Monte Carlo tests at each noise level. The Y axis represented the percentage that the maximum of the wavelet coefficients fall in the interval, which may be regarded as the degree of the detection reliability in this study.

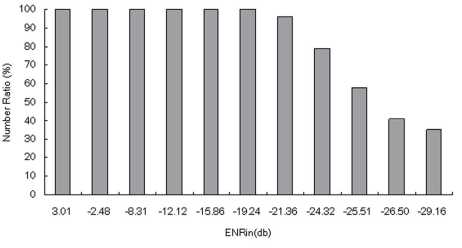


Fig. 5. Statistics results of 100 Monte Carlo tests at different noise levels

Figure 6 showed the mean output ENR of the WT signal detector plotted as a function of the mean input ENR of 100 Monte Carlo tests. The mini bars in the figure represented the standard deviation. From the Fig. 6 we can see that the enhancement of ENR was rather marked even for very low echo-noise-ratios. It should be noticed that the ENR_{out} was positive (larger than 0 db) even in very low ENR_{in}, that is, the signal had a larger amplitude than the noise and can then be discriminated from the noises.

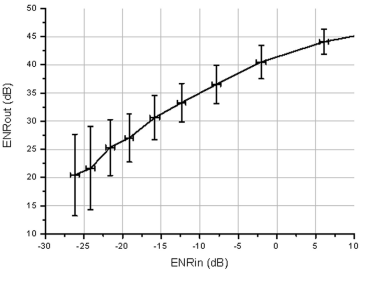


Fig. 6. The mean ENR_{out} for the typical echo in Fig .4 plotted as a function of ENR_{in}

A few samples of the echoes measured in our experiment were shown in Fig. 7. Although the difference between each echo of individual microbubbles did exist and sometimes even was large, the detection based on WT still worked well, as showed in Fig. 8, where 100 individual microbubbles were measured. The notable improvement characterized the large dispersion among the results of the 100 given measurements, which was consist with the diversification of the echoes of individual microbubbles in Fig. 7.

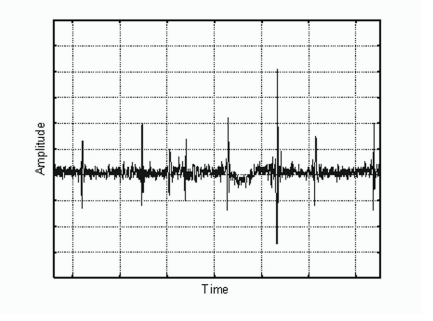


Fig. 7. Some samples of the experimental echoes

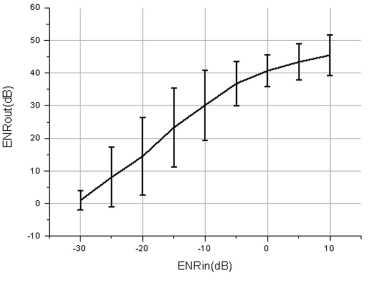


Fig. 8. The mean ENR_{out} of 100 individual microbubbles for different ENR_{in}

4 Discussion

By exploiting the WT coefficient line at S_{max} (the bottom in Fig. 4), the spatial location was showed very close agreement with the experimental echo even for very low ENR_{in}, that is, the spatial solution did not degrade and seemed only dependent on the hardware of the experiment system, hence the trade-off between the wideband spectrum of a insonation pulse and the temporal resolution inherently in the application of harmonic detection was avoided. It should be pointed out that the well-known uncertainty principle still was complied if we reviewed the middle subplot in Fig. 4, where the time resolution (X axis) was good while the frequency resolution (Y axis) degraded. Another thing which deserved mentioned was that the inverse wavelets transform (IWT) was

unnecessary since the detection can be accomplished in the domain of the WT coefficients.

We used -3dB of the maximum of the WT coefficients as a threshold to suppress the noises and Fig. 5 demonstrated a satisfying result, even where ENR_{in} was as low as -21dB, 98% of the intervals of detected echoes fall in the right locations as showed in the middle subplot. Actually, how to select a proper threshold was one of the interesting and widely studied topics on noise reduction with wavelet theory but beyond the scope of this study. So far as the detection of individual microbubbles was concerned, sometimes only knowing maximum of the WT coefficient was enough, for example, in the application of counting number of microbubbles passing through the region of interest (ROI) such as in Dorset' work [10][11].

An important issue in our method was to construct a bubble wavelet as optimal as possible. Although the modified Herring model was certainly able to predict the echo of a single microbubble accurately, nevertheless in most time the corresponding parameters, such as the initial bubble radius, the shell parameters even or the incident ultrasound pressure, may be impossible to be well told. In fact, it was unpractical to find an ideal bubble wavelet for every microbubble. Therefore, what we used in prediction of the bubble wavelet only was the average values of those parameters and it seemed to work well enough. In Fig. 6 and Fig. 8, the improvements of ENR all were achieved by using a "general" bubble wavelet as showed in Fig. 1. Especially for Fig. 8, which resulted from 100 various bubbles, the mean ENR_{out} still reached about 27.95 dB (although lower than 32.70 dB in Fig. 6, where only one microbubble was involved). In our opinion, it was the unique shape, which captured the basic characteristics of the echo of microbubble that played the most important roles in WT detection.

Moreover, the WT seemed to be able to provide more powerful performance than what we understood only from the point view of correlation analysis based on shape of echo. In some extreme cases, the echo of one microbubble was by no means like the predicted shape by theoretical model, which may be due to the collapse of microbubble or some other reasons, nevertheless, this type of echo could also be detected. We guessed that this may be explained by the capability of singularity detection of wavelet transform as described by Mallat[22]. Even so, it was no doubtable that an optimal detection greatly depended upon the choice of an appropriate bubble wavelet. Using a more delicate bubble wavelet will result in higher amplitude wavelet coefficients and hence a higher ENR value.

Acknowledgements. This work was supported by National Natural Science Foundation of China（No. 30570487) and Key Program of Ministry of Education, China (No. 705046).

References

1. Leighton, T.G.: The Acoustic Bubble. Academic Press, London (1994)
2. Miller, D.L., Williams, A.R., Gross, D.R.: Ultrasonic detection of resonant cavitation bubbles in a flow tube by their second-harmonic emissions. Ultrasonics. 22 (1984) 217-224

3. Leighton, T.G., Ramble, D.G., Phelps, A.D.: The detection of tethered and rising bubbles using multiple acoustic techniques. J. Acoust. Soc. Am. 101 (1997) 2626-2635
4. de Jong, N.: Acoustic properties of ultrasound contrast agents. PhD Dissertation Erasmus University, Rotterdam (1993)
5. Goldberg, B.B, Raichlen, J.S., Forsberg, B.: Ultrasound contrast agents: basic principles and clinical applications. 2nd edn. Martin Dunitz, London (2001)
6. de Jong, N., Frinking, P.A., Bouakaz, A., Cate, F.T.: Detection procedures of ultrasound contrast agents. Ultrasonics. 38 (2000) 87–92
7. Unger, E., Matsunaga, T.O., Schumann, P.A., Tutsi, R.: Microbubbles in molecular imaging and therapy. Medicamundi. 47 (2003) 58–65
8. Lamerichs, R., Schäffter, T., Hämisch, Y., Powers, J.: Molecular Imaging: the road to better healthcare. Medicamundi. 47 (2003) 2-9
9. Klibanov, A.L., Rasche, P.T., Hughes, M.S., *et al*.: Detection of Individual Microbubbles of an Ultrasound Contrast Agent: Fundamental and Pulse Inversion Imaging. Acad. Radiol. 9 (2002) S279–S281
10. Droste, D.W., Silling, K., Stypmann, K.J., *et al*.: Contrast Transcranial Doppler Ultrasound in the Detection of Right-to-Left Shunts: Time Window and Threshold in Microbubble Numbers. Stroke. 9 (1999) 1640–1645
11. Droste, D.W., Lakemieier, S., Wichter, T., Stypmann, J., *et al*.: Optimizing the Technique of Contrast Tran-scranial Doppler Ultrasound in the Detection of Right-to-Left Shunts. Stroke. 9 (2002) 2211–2216
12. Zheng, W., Newhouse, V.L.: Onset delay of acoustic second harmonic backscatter from bubbles or microspheres. Ultrasound Med. Biol. 24 (1998) 513—522
13. Shi, W.T., Forsberg, F., Raichlen, J.S., Needleman, L., Goldberg, B.B.: Pressure dependence of subharmonic signals from contrast microbubbles, Ultrasound Med. Biol. 25 (1999) 275-283
14. Bouakaz, A., Frigstad, S., Folkert, J., Cate, T., de Jong, D.: Super harmonic imaging: A new imaging technique for improved contrast detection. Ultrasound Med. Biol. 28 (2002) 59–68
15. Frisch, M., Messer, H.: Detection of a known transient signal of unknown scaling and arrival time. Process IEEE Trans sign process. 42 (1994) 1859-1863
16. Allen, J.S., May, D.J., Ferrara, K.W.: Dynamics of therapeutic ultrasound contrast agents. Ultrasound Med. Biol. 28 (2002) 805-816
17. Morgan, K.E., Allen, J.S., Dayton, P.A., Chomas, J.E., Klibanov, A.L., Ferrara, K.W.: Experimental and theoretical evaluation of microbubble behavior: Effect of transmitted phase and bubble size. IEEE Trans Ultrason Ferroelectr Freq Control. 47 (2000) 1494–1509
18. Dayton, P.A.: The effects of acoustic radiation force on contrast agents: Experimental and theoretical analysis. PhD Dissertation (2001)
19. Daubechies, I.: Ten lectures on wavelets. SIAM, Philadelphia (1992)
20. Du, Y.F., Wan, M.X., Wang S.P., et al.: Surfactant-based nano-shelled microbubble ultrasound contrast agent. Journal of Chemical Industry and Engineering, 54 (2003) 807-812
21. Abbate, A., Koay, J., Frankel, J., Schroeder, S.C., Das, P.: Signal detection and noise suppression using a wavelet transform signal processor: Application to ultrasonic flaw detection. IEEE Trans Ultrason Ferroelectr Freq Control. 44 (1997):14–26
22. Mallat S., Hwang, W.L.: Singularity detection and processing with wavelet. IEEE Trans. Information theory. 38 (1992) 617-638

Using Back Propagation Feedback Neural Networks and Recurrence Quantification Analysis of EEGs Predict Responses to Incision During Anesthesia

Liyu Huang[1,2], Weirong Wang[1,3], and Sekou Singare[2]

[1] Department of Biomedical Engineering, Xidian University, Xi'an, 710071, China
[2] Institute of Biomedical Engineering, Xi'an Jiaotong University, Xi'an, 710049, China
huangly@mail.xjtu.edu.cn
[3] Department of Medical Instrumentation, Shanhaidan Hospital, Xi'an, 710004. China
owl@mail.xidian.edu.cn

Abstract. This paper presents a new approach to detect depth of anaesthesia by using recurrence quantification analysis of electroencephalogram (EEG) and artificial neural network(ANN) . From 98 consenting patient experiments, 98 distinct EEG recordings were collected prior to incision during isoflurane anaesthesia of different levels. The seven measures of recurrence plot were extracted from each of four-channel EEG time series. Prediction was made by means of ANN. Training and testing the ANN used the 'leave-one-out' method. The prediction was tested by monitoring the responses to incision. The system was able to correctly classify purposeful responses in average accuracy of 92.86% of the cases. This method is also computationally fast and acceptable real-time clinical performance was obtained.

1 Introduction

In practice, depth of anesthesia(DOA) is always assessed by anaesthesiologists subjectively by means of observing the blood pressures(BP), heart rate(HR), somatic movement, facial grimacing, lacrimation and diaphoresis. For example, the presence or absence of movement of patients during a surgical procedure can be used to detect level of anaesthesia. The existence of pain can be demonstrated by a rise in blood pressure, tearing, sweating, techycardia, pupillary dilation and so on. The presence of these signs indicates the inadequate level of anaesthesia. However, these clinical signs may be absent due to treatment with muscle relaxants, opioids, cholinergic and β-adrenergic antagonists, vasodilators, and antihypertensive agents. Also, factors like induction agents, patient's illness, age and general health, and site and extent of surgical stimulation, use of controlled ventilation, body temperature, PaCO, and duration of anesthesia may obscure or modify these signs to a large extent making these signs quite unreliable for estimating DOA. For lack of a reliable monitor of anesthetic depth, inadequate levels of anaesthesia occasionally occur, the overall incidence of recall most frequently stated is approximately 1%, cases of intraoperative awareness have been reported in the literature[1][2]. A new non-invasive monitoring technique for detection of DOA would be extremely valuable.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 364 – 373, 2006.

As a target site of anaesthetic effect is the brain, the central nervous system has been the center of attention during anaesthetic course, using electroencephalogram (EEG) to monitor DOA is a reasonable choice. Over the years, numerous efforts have been made by EEG analysis, such as time/frequency domain analysis[3-5], bispectral analysis[6,7], time-frequency distribution (wavelet transform)[8] and so on. However, the use of EEG as a measure of adequacy of anesthesia has achieved limited success. This partly may be attributed to the fact that commonly used signal processing methods is based on the assumption that EEG arises from a linear or stationary process.

Since the middle of the 1980s, some scientists have tried to analyze and study the EEG time series by means of nonlinear theory especially by chaotic dynamics. The fact that EEG possesses chaotic and fractal natures, at least to some extent, has been demonstrated [9-11]. The correlation dimension D_2, which can be used to characterize the complexity of a non-linear dynamical process, has also been used to analyze EEG for evaluating DOA [12].

However, Theiler *et al.*[13, 14] have proved that the brain is nonlinear but is not low dimension chaos. If the EEG is from a high dimension chaotic system, more than 10^m~30^m data are necessary to characterize an m-dimension attractor reliably [15]. So long-time EEG series is needed for reliable prediction, and using D_2 as a measure of DOA becomes impractical. In this paper, a new method is introduced to monitor DOA by using recurrence quantification analysis (RQA) of EEG and artificial neural network(ANN).

2 Materials and Data Acquisition

The Data were obtained from experiments on 98 consenting patients(ASA grade I~II). The patients, 56 men and 42 women, ranging in age from 26 to 63 years(37.8±9.3), were scheduled for elective noncranial surgery.

Anesthesia was induced with propofol, followed by a muscle relaxant to facilitate intubation, and maintained with isoflurane in oxygen. Eelectrocardiogram (ECG), the arterial BP, HR, arterial oxygen saturation, End-tidal CO2(EtCO2, by capnograph) were monitored and all gases(carbon dioxide, isoflurane, oxygen, nitrous oxide) were continuously monitored by mass spectrometer. The arterial pressure and heart rate were measured every minute non-invasively using a Dinamap monitor(Critikon).

Anaesthetic depth has, for the past 30 years, been defined in terms of minimum alveolar concentration(MAC). MAC of an inhalational agent is defined as that alveolar anaesthetic concentration that will prevent movement in 50% of the population subjected to a supramaximal stimulation. This has recently been revised to include intravenous anesthetic agents[16]. A supramaximal stimulation is that stimulation beyond which there is no change in the results of the study[17]. The average MAC value is 1.15% for persons during isoflurane anaesthesia. Responses to skin incision during anesthesia represent a standard test of anaesthetic effect. In our study, we have used this as one of main indicators to assess our prediction.

During and after skin incision, each patient was observed carefully for 2 min to detect purposeful movement, changes in hemodynamic parameters and respiratory pattern.

A gross purposeful movement, usually of the head and limbs, or grimace was considered a positive response, the EEG recording preceding the incision was then labelled as 0.0 for the responder. The response was also estimated as a positive one, in any two of the following three cases, when there were no movement and grimace, but

(1) a significant HR response (rise was greater than 15%);
(2) a significant BP response(rise was greater than 15%);
(3) a spontaneous change in respiratory pattern(change in EtCO2 was more than 12mm Hg).

where the response was negative one, the EEG recording preceding the incision was labelled as 1.0 for non-responder.

Isoflurane concentration must be maintained for a period of time, until the end-tidal concentration of isoflurane stabilized, and a further 15-*min* time period was allowed for equilibration with the brain concentration[18, 19]. The initial surgical incision was made after return of neuromuscular function as determined by a peripheral nerve stimulator and at least 30-*min* after induction of anesthesia, and after at least 10 *min* at a different end-tidal isoflurane concentrations of 1.0MAC(37 patients), 1.2MAC (33 patients) or 1.4MAC(28 patients).

The EEG was monitored using a HXD-I(E) (China) monitor for 2 *min* before incision. Four channels of EEG were recorded using Ag/AgCl electrodes (Fp1, Fp2, O1 and O2, they were placed in accordance with the international standard 10~20 system, with the reference points at left earlobe and right earlobe). The impedance on each of the EEG channel was required to be less than $2K\Omega$ at the start of each experiment. The EEG data were filtered with a high-pass filter at 0.4*Hz* and a low-pass filter at 70*Hz*, a 50*Hz* notch filter was also employed, and sampled at 250*Hz*, digitized to 12bits. From 98 patient experiments, 98 distinct EEG recordings were collected prior to incision during different isoflurane anaesthesia levels, including 31 responder and 67 non-responder recordings. Each recording was 2 *min* long (30,000 points).

3 Method

The proposed approach includes three different procedures. First, RPs are computed from each of four-channel EEG time series, and then the complexity measures are extracted from every RP map. At last, prediction was made by ANN. The input to the ANN will be the 28 measure values (7 from one-channel EEG), the output will be result of the prediction. We describe these procedures briefly as follows.

3.1 Recurrence Plot

Suppose that a dynamic system can be relaxed to an attractor which has an invariant probability distribution in a phase space. The phase space can be constructed by expanding a one-dimensional time series u_i (e.g. EEG, from observations) into a vector set using the Taken's time delay theorem[20],

$$\vec{x}_i = (u_i, u_{i+\tau}, \cdots, u_{i+(m-1)\tau}) . \tag{1}$$

the dimension *m* can be estimated with the method of false nearest neighbors [21, 22], the time delay τ can be estimated with the method of mutual information[23].

Based on the reconstruction of phase space mentioned above, Eckmann *et al.* (1987) first proposed a new method, recurrence plot (RP), to visualize the time dependent behavior of the dynamics of systems, which can be pictured as a trajectory $\vec{x}_i \in \Re^n (i = 1, \cdots, N)$ in the *n*-dimensional phase space[24]. Mathematically speaking, RP can be obtained by the calculation of the $N \times N$ matrix,

$$R_{i,j} := \Theta(\varepsilon_i - \left\| \vec{x}_i - \vec{x}_j \right\|), \qquad i, j = 1, \cdots, N . \tag{2}$$

where $\left\| \cdot \right\|$ is a norm (e.g., the Euclidean norm), ε_i is a cutoff distance, and $\Theta(x)$ is the Heaviside function($\Theta(x) = 1$, if $x > 0$; $\Theta(x) = 0$, if $x \le 0$). In this paper ε_i can be determined by 5% of the maximum value of all norms and the Euclidean norm are used. The binary values in $R_{i,j}$ can be simply visualized by a matrix plot with the colors black (1) and white(0).

3.2 Recurrence Quantification Analysis

Recurrence plots contain subtle patterns that are not easily ascertained by qualitative visual inspection. Zbilut and Webber have presented the recurrence quantification analysis (RQA) to quantify an RP [25, 26, 27], and Marwan et al. extend the definition on the vertical structures[28]. Seven RQA variables are usually examined: (1) *Recurrence rate(RR)*, quantifies a percentage of the plot occupied by recurrent points; (2) *Determinism (DET)*, quantified a percentage between the recurrent points that form upward diagonal line segments and the entire set of recurrence points. The diagonal line consists of two or more points that are diagonally adjacent with no intervening white space; (3)*Laminarity(LAM)*, quantified a percentage between the recurrence points forming the vertical line structures and the entire set of recurrence points; (4)*L-Mean*: quantified an average of the lengthes of upward diagonal lines, this measure contains information about the amount and the length of the diagonal structures in the RP; (5)*Trapping time(TT)*: quantified an average of the lengthes of vertical line structures; (6)*L-Entropy(L-ENTR)*, is from Shannon's information theory, quantified a Shannon entropy of diagonal line segment distributions; (7)*V-Entropy (V-ENTR)*, quantified a Shannon entropy of vertical line segment distributions.

3.3 Artificial Neural Networks

In our study, the input feature vector of the ANN consists of 28 measure values, extracted from each of the four-channel corresponding EEG recordings with label 1.0 (for non-responder) or 0.0 (for responder).

Compared with three-layer(one hidden layer) ANNs, the four-layer ANN(two hidden layers) has a certain advantage in estimating DOA(see Table 2). The number of the second hidden units and the optimum number of clusters are determined according to analysis of input and output feature space[29, 30] and pseudo F-statistic(PFS) clustering technique[31]. The optimum result of ANN structure is 28-8-2-1. We build up the network in such a way that each layer is fully connected to the next layer.

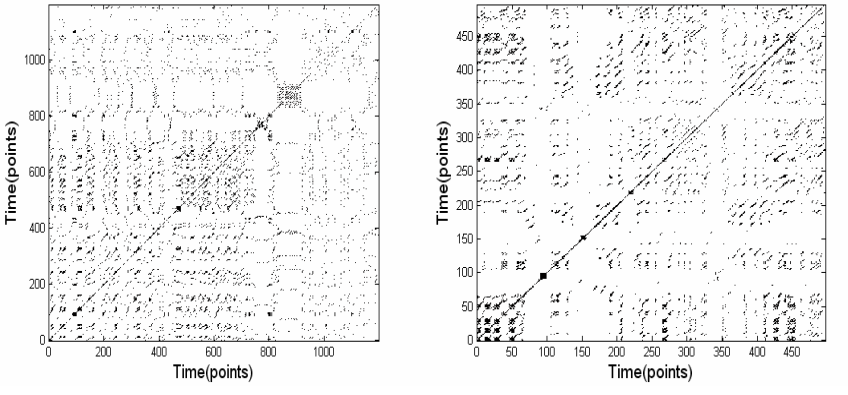
Training of the ANN, essentially an adjustment of the weights, was carried out on the training set, using the back-propagation algorithm. This iterative gradient algorithm is designed to minimize the root mean squared error between the actual output and the desired output.

We trained and tested the ANN using 'leave-one-out' strategy. That is, to train the ANN on samples from *n–1* rats and test on samples from the remaining one. This process is repeated *n* times and each time a different rat is left behind [32].

4 Results

The RP of the EEG recorded was computed and mapped for each patient with a time series of 512 points. Fig 1(a)~(c) show the RPs of three responders with 1.0, 1.2, 1.4 MAC, respectively. (d)~(f) show the RPs of three non-responders with 1.0, 1.2, 1.4 MAC, respectively.

The results of testing our proposed system are shown in Table 1. In total, three response states are misclassified as non-response, and four non-response states are misclassified as response. For response prediction, the average accuracy is 92.86% (see Table 2). For different prediction schemes, the comparison of performances is shown in Table 2.



(a) (b)

Fig. 1. RP of typical patients, including three responders with 1.0 MAC (a), 1.2 MAC (b), 1.4 MAC (c), and three non-responders with 1.0 MAC (d), 1.2 MAC (e), 1.4 MAC (f)

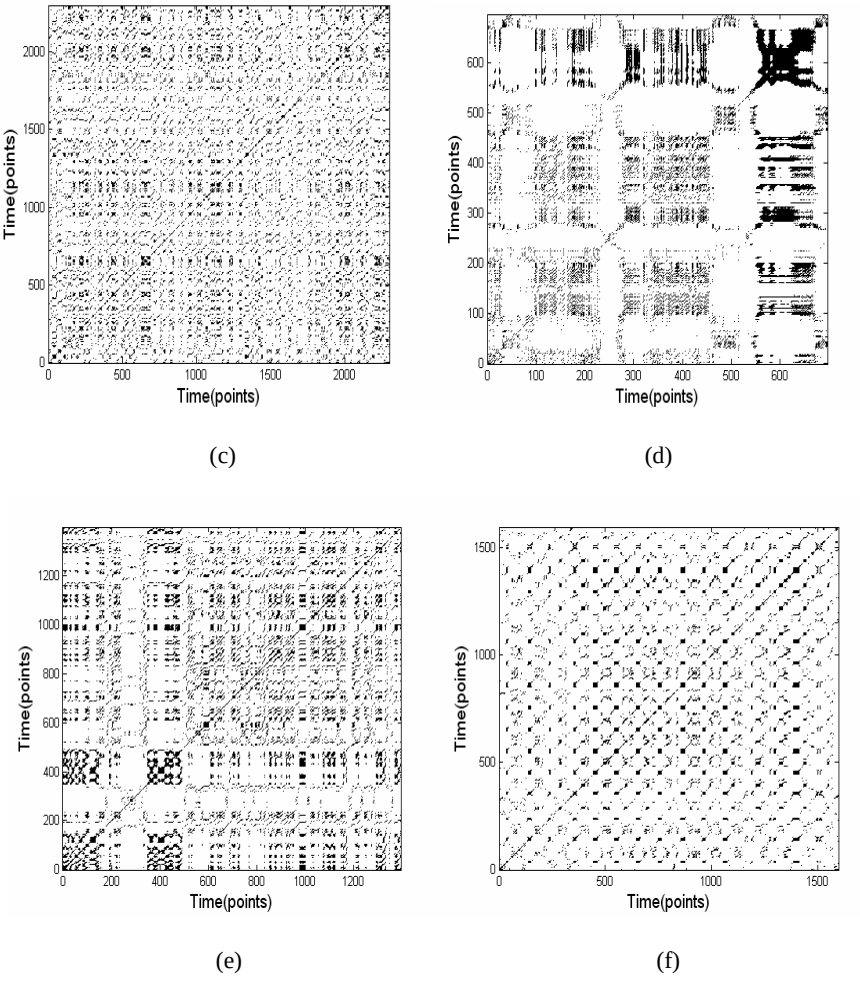


Fig. 1. (*Continued*)

We tested our system using different ANNs, for four-layer (28-8-2-1) and three-layer (28-8-1) ANNs, the accuracy is 92.86% and 87.76%, respectively.

It would be informative to briefly compare our system with the commonly used spectral analysis methods, such as the bispectral analysis (BIS)[6, 7], the spectral edge frequency (SEF) and the median frequency (MF) [33, 34].

The accuracy for BIS, SEF and MF methods is 84%, 79% and 68%, respectively[7]. Another accuracy 83% was given out by another researcher, Sebel[35], also using bispectrum method. Some of the results are cited from the references (paper [7] and [35]) for comparison.

Table 1. Testing results for proposed system, which employs four-layer ANN

Level of anesthesia	Size of training set	Number of patients	Misclassified rate for response state	Misclassified rate for non-response state	Accuracy
1.0MAC	97	37	3/21	0/16	91.89%
1.2MAC	97	33	0/7	3/26	90.91%
1.4MAC	97	28	0/3	1/25	96.43%

Table 2. Comparison of performances of different prediction schemes

Predicting schemes	Other conditions	Types of ANN employed	Average accuracy
RQA using ANN	98 patients and 98 EEG recordings at different anesthetic levels	Four-layer ANN (28-8-2-1)	92.86%
		Three-layer ANN (28-8-1)	87.76%
Spectral analysis[a]	BIS		83%
Spectral analysis[a]	BIS		84%
	SEF		79%
	MF		68%

[a] These results are cited from the references, paper [7] and [35], respectively, compared with our conclusion.

5 Discussion

The study of EEG in monitoring the DOA has been cited as early as 1940[3], however, the use of EEG as a measure of adequacy of anesthesia has achieved limited success. The passed studies show that adequacy of anesthesia is a complicated concept and it is difficult to accurately evaluate DOA by a single parameter[36]. The recurrence plot represents the recurrence of the phase space trajectory to a certain state, which is a fundamental property of deterministic dynamical systems[37, 38]. The chaotic or quasi-random behavior of EEG signals gives us the possibility to quantitatively study functional changes of the brain by means of RQA. In this paper, we propose a new approach to predict response to incision during anaesthesia using recurrence quantification analysis method. Compared with other schemes, our designed system has a better performance and the RQA is a potential way to assess the anaesthetic adequacy.

The anesthetic dose versus the depth of anesthesia curve is highly nonlinear [39] which motivated the use of ANN's as the basic classifiers in our design. Owing to the restricted experimental conditions, we have only 98 recordings from 98 patients for training and testing ANN. A good way to tackle such a dilemma would be to train and test ANN using 'leave-one-out' method [32]. It is a standard method to evaluating classification systems, especially in the case of small samples.

The designed system consists of two operations: off-line training of the ANN and on-line operation of prediction of responses during anaesthesia. These two operations contain similar function blocks: EEG collection, RP and its measure extraction and four-layer ANN for prediction. Before the system goes into operation, a specific database must be built for off-line training of the ANN. After training, the weights are fixed and can be used for on-line work. The last decision is made according to the output of the ANN. During the on-line application, the recorded EEG and other parameters can also be stored in the specific database for updating. Thus, every certain period, the ANN is retrained off-line using the newly updated specific database, and then the new weights are sent to the trained ANN to update its weights.

The RP and its measure analysis are much more computationally efficient than conventional analysis methods, such as FFT and D2. Moreover, after off-line training, the ANN can implement on-line forward propagation at high speed. In our experiment, the average time for computing RPs and their measures of EEGs, then prediction by ANN is about 0.057s on our computer (Toshiba Satellite Pro 4600 with Pentium III 1G CPU). So our system can be computationally enough fast and suited for real-time clinical application.

Nevertheless, the work reported here is preliminary. More studies need to be performed to test the effectiveness of the system, especially with different inhalational anaesthetic agents and with opioid anaesthesia, where awareness during surgery is a major concern.

Acknowledgment

This work was supported by China Postdoctoral Science Foundation and National Natural Science Foundation of China under grant NO. 60371023.

References

1. Moerman, N., Bonke, B., Oosting, J.: Awareness and recall during general anesthesia: facts and feelings. Anesthesiology, 79 (1993) 454–464
2. Ghoneim, M.M., Block, R.L.: Learning and consciousness during general anaesthesia. Anesthesiology, 76(1992) 279–305
3. Rubin, M.A., Freeman, H.: Brain potential changes in man during cyclopropane anesthesia. J Neurophysiol, 3(1940) 33–42
4. Thomsen, C.E., Christensen, K.N., Rosenflack, A.: Computerized monitoring of depth of anaesthesia with isoflurane, Br. J. Anaesthesia. 63(1989) 36–43.
5. Sharma, A., Roy, R.J.: Design of a recognition system to predict movement during anaesthesia, IEEE Trans. on Biomed. Eng., 44(1997) 505–511.

6. Kearse, L. A., Manberg, P., DeBros, F., (eds.): Bispectral analysis of the electroencephalogram during induction of anesthesia may predict hemodynamic responses to laryngoscopy and intubation. Electroencephalography and clinical Neurophysiology, 90(1994) 194–200
7. Vernon, J. M., Lang, E., Sebel, P. S., (eds.): Prediction of movement using bispectral electroencephalographic analysis during propofol/alfentanil or isoflurane/alfentanil anes- thesia. Anesth Analg, 80(1995) 780–785
8. Nayak, A., Roy, R. J., Sharma, A.: Time-frequency spectral representation of the EEG as an aid in the detection of depth of anaesthesia, Ann. Biomed. Eng., 22(1994) 501–513.
9. Pradhan, N., Adasivan, P.K.: Validity of dimensional complexity measures of EEG signals, Int. J. Bifurcation Chaos Appl. Sci. Eng., 7(1997) 173–286.
10. Ogo, K., Nakagawa, W.: Chaos and fractal properties in EEG data, Electron. Commun.Jpn., 78(1995) 27–36
11. Yaylali, I., Kocak, H., Jayakar, P.: Detection of seizures from small samples using nonlinear dynamic system theory. IEEE Trans. on Biomed. Eng., 43(1996) 743–751.
12. Jackson, M.E., Cauller, L.J.: Non-linear dynamics of neocortical spontaneous field potentials during anesthetized and awake states in chronically implanted rats, Soc. Neurosci.Abst., 21(1995) 57.10
13. Theiler, J.: Testing for nonlinearity in time series: the method of surrogate data. Physica D, 58(1992) 77–94
14. Pritchard, W. S.: Dimensional analysis of resting human EEG II: Surrogate data testingindicates nonlinearity but not low-dimensional chaos. Psychophysiology, 32(1995) 486–491
15. Wolf, A., Swift, J. B., Swinney, H. L., (eds.): Determining Lyapunov exponents from atime series. Physica D, 16(1985) 285–217
16. Buhrer, M.: Thiopental pharmacodynamics. Anesthesiology, 77(1992) 226–23617
17. Zbinden, A. M., Maggiorini, M., Petersen-Felix, S., (eds.): Anesthetic depth defined using multiple noxious stimuli during isoflurane/oxygen anesthesia. Anesthesiology, 80(1994) 253–260
18. Steward, A., Allott, P. R., Mapleson, W. W.: The solubility of halothane in canine blood and tissues. Br J Anaesthesia, 47(1975) 423–433
19. Mapleson, W. W.: Circulation-time models of the uptake of inhaled anaesthetics and data for quantifying them. Br J Anaesthesia, 45(1973) 319–333
20. Takens, F.: Determining strange attractors in turbulence, Lecture notes in math., 898(1981) 361–381.
21. Cao, L.: Practical method for determining the minimum embedding dimension of a scalar time series, Physica D, 110(1997) 43–50.
22. Kennel, M.B.: Determining embedding dimension for phase-space reconstruction using a geometrical construction, Physical Review A, 65(1992) 3403–3411.
23. Fraser, A. M., Swinney, H. L.: Independent coordinates for strange attractors from mutual information. Phys Rev A, 33(1986) 1134–1140
24. Eckmann, J. P., Kamphorst, S. O., Ruelle, D.: Recurrence plots of dynamical systems.Europhys. Lett. 4(1987) 973–977.
25. Webber, Jr. C.L., Zbilut, J.P.: Dynamical assessment of physiological systems and states using recurrence plot strategies. J. Appl. Physiol., 76(1994) 965–973.
26. Zbilut, J. P. and Webber, Jr. C.L.: Embeddings and delays as derived from quantification of recurrence plots. Phys. Lett. A, 171(1992) 199–203.
27. Trulla, L.L., Giuliani, A., Zbilut, J. P., Webber, Jr. C. L.: Recurrence quantification analysis of the logistic equation with transients. Phys. Lett. A, 223(1996) 255–260.

28. Marwan, N., Wessel, N., Meyerfeldt, U., (eds.): Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. Phys. Reviw. E, 66(2002) 026702
29. Lippmann, R.P.: An Introduction to Computing with Neural nets. IEEE ASSP Magazine, (1987 Apr.) 4–22
30. Mirchandami, G., Cao, W.: On hidden nodes for neural nets. IEEE Trans on Circuits and System, 36(1989) 661–664
31. Vogel, M. A., Wong, A. K. C.: PFS clustering method. IEEE Trans on Pattern Anal Mach Intell, 1(1979) 237–245
32. Fukunaga, K.: Introduction to Statistical Pattern Recognition. 2nd edition, Academic, San Diego, CA (1990)
33. Drummond, J. C., Brann, C. A., Perkins, D. E., (eds.): A comparison of median frequency, spectral edge frequency, a frequency band power ratio, total power, and dominance shift in determination of depth of anaesthesia. Acta Anaesthesiologica Scand., 35(1991) 693–699
34. Rampil, I. J., Matteo, R. S.: Changes in EEG spectral edge frequency correlate with the hemodynamic response to laryngoscopy and intubation. Anesthesiology, 67(1987) 139–142
35. Sebel, P.S., Bowles, S.M., Saini, V., (eds.): EEG bispectrum predicts movement during thiopental/isoflurane anaesthesia. J Clin Monit, 11(1995) 83–91
36. Linkens, D. A.: Adaptive and intelligent control in anesthesia. IEEE Contr. Syst. Technol., (1992 Dec.) 6–11
37. Argyris, J.H., Faust, G., Haase, M.: An Exploration of Chaos, North-Holland, Amsterdam (1994)
38. Ott, E.: Chaos in Dynamical Systems, Cambridge University Press, Cambridge (1993)
39. Quasha, A. L., Eger, E. I., Tinker, H. H.: Determination and applications of MAC, Anesthesiol., 53(1980) 315–334

Numerical Simulations of Contribution of Chemical Shift in Novel Magnetic Resonance Imaging

Huijun Sun, Tao Lin, Shuhui Cai, and Zhong Chen*

Department of Physics, The School of Physics and Mechanical & Electrical Engineering, Xiamen University, Xiamen, Fujian 361005, P.R. China
Tel.: +86-592-2183717; Fax: +86 0592 218 9468
chenz@xmu.edu.cn
http://pmee.xmu.edu.cn/index.htm

Abstract. Contribution of chemical shift to intermolecular multiple-quantum coherence (iMQC) imaging signals in two-component systems was simulated and discussed using an efficient numerical algorithm based on the Bloch equations with an additional nonlinear term describing distant dipolar field. Numerical simulation switches back and forth between real and Fourier spaces to handle dipolar field effects in three-dimensional sample. The iMQC signals of each component of two-component systems can be obtained respectively when the second pulse of the CRAZED pulse sequence is selective. Simulation results show that chemical shift provides an edge detection method to regions containing spins with chemical shift offset and selected by the second RF pulse, and different gray value is related to different chemical shift in detected regions. These results indicate that chemical shift may provide new imaging information helpful for iMQC magnetic resonance imaging.

1 Introduction

In conventional magnetic resonance imaging (MRI), chemical shift in some instances introduces artifacts, such as chemical shift mismatch artifacts (CSMAS) [1], which result in a blur boundary or ghost image. This is similar to the effects of different magnetic susceptibilities in MRI, which cause the variation of imaging signal intensity at the edges of metallic implant, tumor and bone-tissue [2,3]. However, these bad effects may become favourable for object detection in the intermolecular multiple-quantum coherence (iMQC) MRI. In addition, chemical shift imaging (CSI) is an established tool in vivo biochemistry study and fundamental biomedical research. CSI is also emerging as a valuable diagnostic method in clinic. It has been successfully applied to detecting at a very early stage tumor response to chemotherapy [1], or to studying brain metabolic alterations in patients with AIDS [1]. Therefore, the study and applications of chemical shift in MRI have become increasing important and valuable.

In recent years iMQC has been a hot topic in nuclear magnetic resonance comm.-unity. Warren and coworkers first achieved successfully intermolecular zero-quantum

* Corresponding author.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 374 – 383, 2006.

coherence (iZQC) imaging of in vivo rat brain in 1998 [4]. Zhong and Chen et al. obtained the first multislice human brain images of intermolecular double-quantum coherence (iDQC) on a clinical 1.5T MR scanner in 2000 [5]. In addition, iMQC has also been studied and applied deeply in functional magnetic resonance imaging (fMRI) [6]. Although the signal of iMQC has lower intensity than conventional one, it includes specific imaging parameters such as dipolar correlation distance, providing new contrast mechanism for imaging, and can get higher contract and resolution. All these give iMQC a seductive foreground in MRI.

Warren's group has simulated contribution of chemical shift in iMQC MRI using the CRAZED pulse sequence with two hard pulses [7,8]. In this paper, contribution of chemical shift to iZQC and iDQC imaging signals of two-component systems was simulated and discussed using a modified CRAZED pulse sequence, where the second pulse was selective.

2 Simulation Algorithm

Fig. 1 shows the modified CRAZED pulse sequence, where the second radio-frequency (RF) pulse β is selective for either I or S spins in a two-component system [9,10]. Imaging signals were acquired at maximum intensity of spin echo in simulation, so the frequency encoding and phase encoding gradients are omitted.

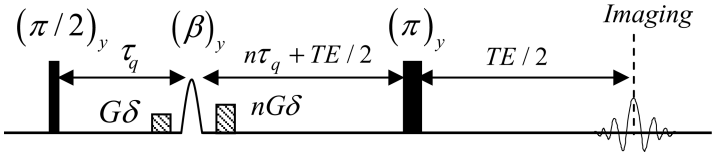


Fig. 1. The pulse sequence for investigating the iMQC signal behaviors in a two-component system without J coupling. The second RF pulse is frequency selective.

Theoretically, the signals for coherence orders n = 0 and –2 have strongest intensities among all coherence orders, and the signal intensities of iZQC (n = 0) and iDQC (n = –2) of two components reach maximum when the sequence parameters are optimal [10]. To obtain maximal signals of I spin, the second RF pulse β should be phase cycled between 45° and 135° for n = 0, and be 60° for n = –2. For the signals of S spin β is 90°.

To achieve a good quantitative description of iMQC signals in MRI, we use a two-component system consisting of I and S spins without J coupling as an example. Let Ω_I and Ω_S be the resonance frequency offsets of I and S spins in the rotating frame, T_1^{I} and T_1^{S} be the single-quantum coherence (SQC) longitudinal relaxation times, and T_2^{I} and T_2^{S} be the SQC transverse relaxation times of I and S spins, respectively. Ignoring other factors, and only considering chemical shift and relaxation, the final signal of I spin can be expressed as

$$M_{obs}^{I}(t_2) = -\frac{3\sqrt{3}i^{-1}}{16}(\gamma_0\mu_0 M_0^I t_2)e^{i(\tau_q+t_2)\Omega_I} M_0^I e^{-(2\tau_q+t_2)/T_2^I} \qquad n=-2$$
$$M_{obs}^{I}(t_2) = -\frac{i}{4}(\gamma_0\mu_0 M_0^I t_2)e^{i(\tau_q+t_2)\Omega_I} M_0^I e^{-(2\tau_q+t_2)/T_2^I} \qquad n=0 \tag{1}$$

and the final signal of S spin is

$$M_{obs}^{S}(t_2) = \frac{1}{3}(\gamma_0\mu_0 M_0^I t_2)M_0^S e^{i(\Omega_I+\Omega_S)(\tau_q+t_2)} e^{-(\tau_q+t_2)/T_2^S-\tau_q/T_2^I} \qquad n=-2,0 \tag{2}$$

where γ is the gyromagnetic ratio, μ_0 is the magnetic vacuum permeability, M_0 represent the equilibrium magnetization of spins per unit volume, t_2 is detection time.

2.1 Simulation Method

Assume the correlation gradients are along z direction, according to the classical theory, the time evolution of spin magnetizations can be described by non-linear Bloch equation

$$\frac{\partial \mathbf{M}(\mathbf{r},t)}{\partial t} = \gamma \mathbf{M}(\mathbf{r},t)\times\left\{\left[\frac{\Omega(\mathbf{r})}{\gamma}\hat{\boldsymbol{z}} + G(\mathbf{r},t)z\right] + \mathbf{B}_r(\mathbf{r},t) + \mathbf{B}_d(\mathbf{r},t)\right\} + D\nabla^2\mathbf{M}(\mathbf{r},t) - \frac{M_x(\mathbf{r},t)\hat{\boldsymbol{x}} + M_y(\mathbf{r},t)\hat{\boldsymbol{y}}}{T_2(\mathbf{r})} + \frac{M_0(\mathbf{r}) - M_z(\mathbf{r},t)}{T_1(\mathbf{r})}\hat{\boldsymbol{z}} \tag{3}$$

where D is the translational self-diffusion coefficient of spins; $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{z}}$ are unit vectors along the three orthogonal coordinate axes. The simulation process is solving Eq. (3) on each time step under pulse sequence by using a Runge-Kutta algorithm, starting from initial parameters and conditions. In the equation, the radiation damping term $\mathbf{B}_r(\mathbf{r},t)$ is easy to calculate in real space if RF pulse is uniform over the whole sample, but diffusion term $D\nabla^2\mathbf{M}(\mathbf{r},t)$ and distant dipolar field term $\mathbf{B}_d(\mathbf{r},t)$ are difficult to calculate in real space. The expression for the distant dipolar field is

$$\mathbf{B}_d(\mathbf{r}) = \frac{\mu_0}{4\pi}\int d^3\mathbf{r}' \frac{1-3\cos^2\theta_{rr'}}{2|\mathbf{r}-\mathbf{r}'|^3}[3M_z(\mathbf{r}')\hat{\boldsymbol{z}} - \mathbf{M}(\mathbf{r}')] \tag{4}$$

where $\theta_{rr'}$ is the angle between the inter-nuclear vector $\mathbf{r}–\mathbf{r}'$ and the main magnetic field. Since $\mathbf{B}_d(\mathbf{r})$ must be integrated over the magnetization of the whole sample, it is quite computationally expensive to calculate the dipolar field from Eq. (4) directly. Fortunately, this equation can be simplified in k-space,

$$\mathbf{B}_d(\mathbf{k}) = \frac{\mu_0}{6}[3(\hat{\mathbf{k}}\bullet\hat{\boldsymbol{z}})^2 - 1][3M_z(\mathbf{k})\hat{\boldsymbol{z}} - \mathbf{M}(\mathbf{k})] \tag{5}$$

Similarly, the calculation of the diffusion operator $D\nabla^2$ is reduced to multiplication $-Dk^2$ in Fourier space [7,8]. Fig. 2 gives a schematic of the numerical algorithm.

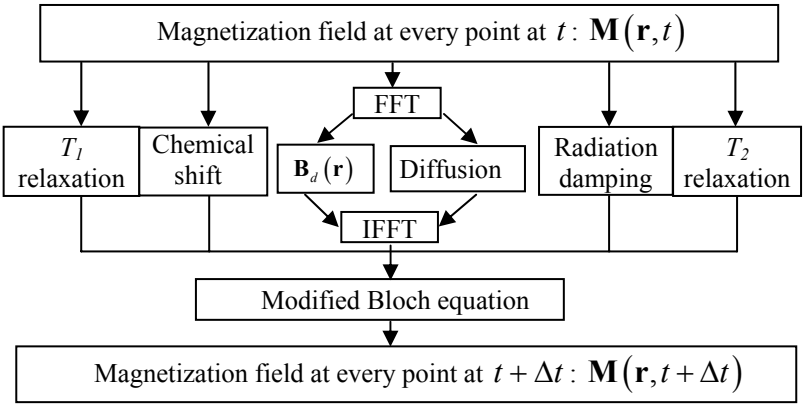


Fig. 2. Schematic of the algorithm for integration of the non-linear Bloch equation

For two-component systems, each spin evolves simultaneously according to individual Bloch equation. Chemical shift and relaxation effects of each spin are independent respectively, while distant dipolar field, diffusion, and radiation damping of each spin are interactive with each other. So in simulation we replaced distant dipolar field term of each spin with interactive term, such as $\boldsymbol{B}_d^I \rightarrow (\boldsymbol{B}_d^I + \boldsymbol{B}_d^S)$ and $\boldsymbol{B}_d^S \rightarrow (\boldsymbol{B}_d^I + \boldsymbol{B}_d^S)$.

The frequency encoding gradient and phase encoding gradients were not considered in simulation, because they were only used for spatial location of imaging information. Imaging signals were sampled at the center of spin echo (maximum intensity), and calculated with orthogonal detection, namely $M_{abs}=abs(M_x+iM_y)$. Two-dimensional images came from the final signals summed along z-direction, along which gradient pulses were applied.

2.2 Initial Parameters for Imaging Sample

Two homo-nuclear components (I and S spins) with different chemical shift offsets were assumed to mix uniformly. Each component had its own initial parameter set. The sample used for simulation is shown in Fig. 3. It is a cube (solid) with a length of 2.8 mm (56×56×56 grid points along x, y, and z directions). To prevent anomalous edge effects, the sample must be surrounded by empty space. Therefore, the sample is zero-padded by 4 grid points at the edges along each direction, resulting in a cube box (dashed) with a length of 3.2 mm and 64×64×64 grid points. The shaded cube (24×24×24 grid points) with a length of 1.2 mm at the center of the sample has slightly different chemical shift from the other region of the sample.

The evolution time (τ_q) and echo time (TE) were set to 10 and 100 ms, respectively. The amplitude of the gradient pulse G was 5.87 G/cm and the duration δ was 1 ms. The correlation distance controlled by the gradient pulse was then calculated to be 200 μm. Other parameters of each component are listed in Table 1.

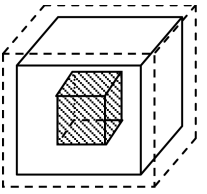


Fig. 3. Sample used for simulation. The region of simulation is the dashed box, the solid cube is the sample region, and the shaded box has chemical shift offsets.

Table 1. Parameters for each spin

Spin	M_0 (A/m)	Ω outside the shaded region (Hz)	T_1 (s)	T_2 (ms)
I	0.04	0	1.5	150
S	0.02	200	1.0	100

3 Results and Discussion

Ignoring diffusion and radiation damping effects, we simulated contributions of chemical shift to iZQC and iDQC imaging signals of two components for two cases: (1) I spin in the shaded region had chemical shift offsets relative to I spin in the other region while the chemical shift of S spin was uniform in the whole sample; (2) S spin in the shaded region had chemical shift offsets relative to S spin in the other region and the chemical shift of I spin was uniform in the whole sample. The second RF pulse was optimized to attain maximal iZQC and iDQC signals of I and S spins.

3.1 Imaging Signals of I and S Spins in the First Case

A. Imaging Signals of I Spin

Figure 4 shows the simulation results for coherence order n = –2 of I spin when the chemical shift offset of I spin in the shaded region was –50 Hz.

Since the selective RF pulse is only applied on I spin, the iMQC signals of I spin are not contributed by S spin. Based on Eq. (1), the variation of the signals with chemical shift for n = –2 and n = 0 can be written as $M_{obs}^{I} \propto e^{2\Omega_I(\tau_q+t_2)}$ when the chemical shift of I spin is uniform. If the chemical shift offset of I spin in the shaded region is notated as Ω_I', the signal intensity then becomes $M_{obs}^{I} \propto e^{(\Omega_I+\Omega_I')(\tau_q+t_2)}$ at the edge of the shaded region. Because the π RF pulse in detection period cannot completely remove the effect of chemical shift in the evolution period at the edge of the shaded region, the signals there are not completely refocused. In addition, the total phase evolution after a time interval of $2\tau_q$ is $\pm\tau_q(\Omega_I-\Omega_I')$, which is not influenced

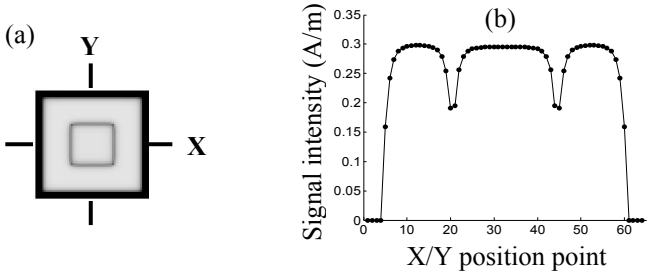


Fig. 4. The simulation results for coherence order n = –2 of I spin. The chemical shift offset of I spin in the shaded region was –50Hz. (a) Imaging result. (b) The signal intensity distribution of the center slice along x or y direction in the imaging result.

by the π RF pulse. Therefore, the signal intensity at the edge of the shaded region decreases as $\cos[\tau_q(\Omega_I - \Omega_I')]$. These two reasons result in attenuation of the signal intensity at the edge of the shaded region. These results agree with the simulation results using the CRAZED sequence with non-selective RF pulse [7].

B. Imaging Signals of S Spin

Figure 5 shows the simulation results for coherence order n = –2 of S spin when the chemical shift offset of I spin in the shaded region was –50Hz. The Simulation results show that the signal intensity of S spin not only attenuates at the edge of shaded region, but also lowers in the whole region.

Since the selective β pulse is not applied on S spin, the iMQC signals of S spin result from the distant dipolar field caused by I spin. Based on Eq. (2), the variation of the signals with chemical shift for n = –2 and n = 0 can be written as

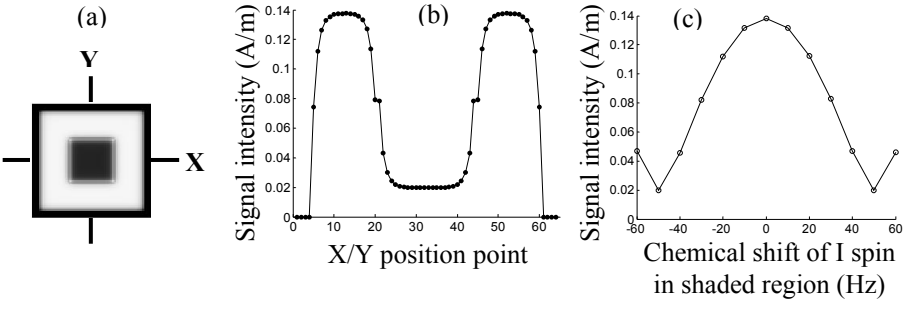


Fig. 5. The simulation results for coherence order n = –2 of S spin. (a,b) The chemical shift offset of I spin in the shaded region was –50 Hz. (a) Imaging result. (b) The signal intensity distribution of the center slice along x or y direction in the imaging result. (c) The signal intensity at the center point of the image of S spin when the chemical shift offsets of I spin in the shaded region varied.

$M_{obs}^{S} \propto e^{(\Omega_I+\Omega_S)(\tau_q+t_2)}$. When the chemical shift of I spin is uniform, the π RF pulse in the detection period removes the effect of chemical shift Ω_I and Ω_S. When the chemical shift offset of I spin in the shaded region is Ω_I', M_{obs}^{S} in the shaded region is different from the outside one. At the edge of the shade region, the signal intensity decreased similarly to I spin.

We also simulated the imaging signals of S spin with Ω_I' varying from –60 Hz to 60 Hz in an increment of 10 Hz. Fig. 5c shows the signal intensity at the center point of image. The variation of the intensity approximately obeys the relation of $\cos[\tau_q(\Omega_I-\Omega_I')]$. So does the signal intensity at the other point of the shaded region.

3.2 Imaging Signals of I and S Spins in the Second Case

A. Imaging Signals of I Spin

According to the theory, the iMQC signals of I spin are only contributed by I spin, not by S spin. Therefore, when S spin in the shaded region has chemical shift offsets different from the other region and the chemical shift offset of I spin is uniform in the whole sample, the imaging signal intensity of I spin is uniform. Simulation results agree well with theoretical prediction [10].

B. Imaging Signals of S Spin

Figure 6 shows the simulation results for coherence order n = –2 of S spin when the chemical shift offset Ω_S' in the shaded region was 150 Hz. The signal intensity also lowers in the whole shaded region. At the edge of the region, the signals decreased rapidly due to the uniform chemical shift offset of I spin, evidently different from the first case (Fig. 5b). The signal intensity in the shaded region varies according to $\cos[\tau_q(\Omega_S-\Omega_S')]$ (Fig. 6c).

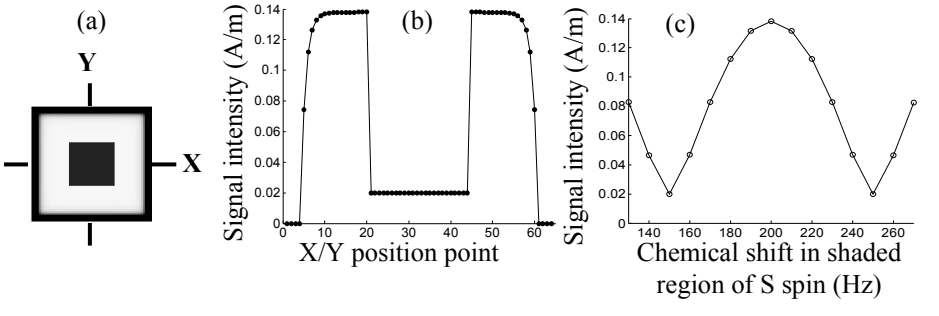


Fig. 6. The simulation results for coherence order n = –2 of S spin. (a,b) The chemical shift offset of S spin in the shaded region was 150 Hz. (a) Imaging result. (b) The signal intensity distribution of the center slice along x or y direction in the imaging result. (c) The signal intensity at the center point of the image of S spin when the chemical shift offsets of S spin in the shaded region varied.

Above we have discussed the iDQC imaging signals of I and S spins. The iZQC imaging signals of I and S spins had identical features with iDQC.

3.3 Imaging Signals of I and S Spins with a Selective π Refocusing Pulse

The effect of chemical shift in evolution period τ_q can be removed when a π refocusing pulse is added at the center of τ_q, as shown in Fig. 7. We simulated imaging signals of I and S spins in the same two cases as above when the π refocusing pulse in τ_q was selective.

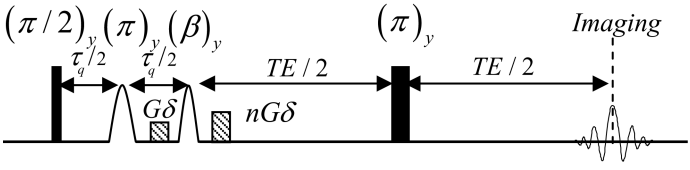


Fig. 7. The pulse sequence with a selective π refocusing pulse in τ_q

Figure 8 shows the simulation results of I and S spins for iZQC when the π refocusing pulse in τ_q excites only I spin. The π pulse refocuses the chemical shift evolution of I spin in τ_q, but not the chemical shift evolution of S spin. As a result, the signal intensity of I spin is almost identical in the whole sample although there is chemical shift variation of I spin in the shade region (see Fig. 8a). A little attenuation at the edge of the shaded region is due to the incompletely refocused magnetization resulting from the limited number of sample points around the edges, which introduces small differences in evolution before and after the selective π refocusing pulse. In Fig. 8b the signal intensity of S spin attenuates in the whole shaded region when there is a chemical shift variation of S spin in the shade region.

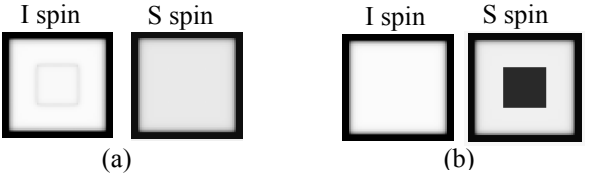


Fig. 8. The simulation results of I and S spins for iZQC when the selective π refocusing pulse in τ_q excites only I spin. (a) $\Omega_I' = -50\text{Hz}$. (b) $\Omega_S' = 150\text{Hz}$.

When the π refocusing pulse in τ_q excites only S spin, the simulation results of I and S spins for iZQC are shown in Fig. 9. This pulse refocuses only the chemical shift evolution of S spin in τ_q. Therefore, the attenuation of the signal intensity of I and S spins retains when there is a chemical shift variation of I spin in the shade region (Fig. 9a), while the signal intensity of I or S spin is identical in the whole sample when the chemical shift variation in the shade region comes from S spin (Fig. 9b).

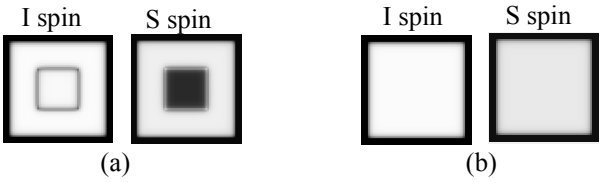


Fig. 9. The simulation results of I and S spins for iZQC when the selective π refocusing pulse in τ_q excites only S spin. (a) $\Omega_I' = -50\text{Hz}$. (b) $\Omega_S' = 150\text{Hz}$.

Similar to reference [7], subtracting the results obtained from the pulse sequence in Fig. 6 from the results obtained from the pulse sequence in Fig. 1 leaves the contribution of chemical shift. The subtraction image is almost specific to areas containing spins with different frequency offsets, and provides signal contrast for these areas.

4 Conclusions

The simulation algorithm based on the non-linear Bloch equation provides a fast, efficient computing method, which can present intuitive, exact imaging results for 3D sample in MRI. Different from conventional MRI, contribution of chemical shift in iMQC imaging can not only offer an edge detection tool in MRI, but also give different gray value of image in detected regions. Our simulation results show that chemical shift in iMQC imaging is a potential parameter for studying MRI. It produces a new interesting image contrast different from conventional one, which may be useful for fMRI and MRI.

Acknowledgments

This work was partially supported by the NNSF of China under Grants 10234070 and 10575085, and the Program for NCET of MOE of China.

References

1. Pohmann, R., von Kienlin, M., Haase, A.: Theoretical Evaluation and Comparison of Fast Chemical Shift Imaging Methods. J. Magn. Reson. 129 (1997) 145-160
2. Pauchard, Y., Smith, M.R., Mintchev, M.P.: Improving Geometric Accuracy in the Presence of Susceptibility Difference Artifacts Produced by Metallic Implants in Magnetic Resonance Imaging. IEEE Trans. Med. Imaging 24 (2005) 1387-1399
3. Fessler, J.A., Lee, S., Olafsson, V.T., Shi, H.R., Noll, D.C.: Toeplitz-Based Iterative Image Reconstruction for MRI with Correction for Magnetic Field Inhomogeneity. IEEE Trans. Signal Process. 53 (2005) 3393-3402
4. Warren, W.S., Ahn, S., Mescher, M., Garwood, M., Ugurbil, K., Richter, W., Rizi, R.R., Hopkins, J., Leigh, J.S.: MR Imaging Contrast Enhancement Based on Intermolecular Zero Quantum Coherences. Science, 281 (1998) 247-251

5. Zhong, J.H., Chen, Z., Kwok, E.: In vivo Intermolecular Double-Quantum Imaging on a Clinical 1.5(7)T MR Scanner. Magn. Reson. Med. 43 (2000) 335-341
6. Zhong, J.H., Chen, Z., Kwok, W.E., Kennedy, S., You, Z.Y.: Optimization of Blood Oxygenation Level-Dependent Sensitivity in Magnetic Resonance Imaging Using Intermolecular Double-Quantum Coherence. J. Magn. Reson. Imaging 16 (2002) 733-740
7. Garrett, R.S., Warren, W.S.: Numerical Studies of Intermolecular Multiple Quantum Coherences: High-Resolution NMR in Inhomogeneous Fields and Contrast Enhancement in MRI. J. Magn. Reson. 146 (2000) 1-13
8. Enss, T., Ahn, S., Warren, W.S.: Visualizing the Dipolar Field in Solution NMR and MR Imaging: Three-Dimensional Structure Simulations. Chem. Phys. Lett. 305 (1999) 101-108
9. Ahn, S., Lisitza, N., Warren, W.S.: Intermolecular Zero-Quantum Coherences of Multi-Component Spin Systems in Solution NMR. J. Magn. Reson. Imaging 133 (1998) 266-272
10. Chen, Z., Chen, Z.W., Zhong, J.H.: Quantitative Characterization of Intermolecular Dipolar Interactions of Two-Component Systems in Solution Nuclear Magnetic Resonance. J. Chem. Phys. 115 (2001) 10769-10779

Secrecy of Signals by Typing in Signal Transduction*

Min Zhang[1,**], Guoqiang Li[2], and Yuxi Fu[1]

[1] BASICS, Department of Computer Science and Engineering,
Shanghai Jiao Tong University, Shanghai 200030, China
{zhangmin, yxfu}@sjtu.edu.cn
[2] Japan Advanced Institute of Science and Technology
Asahidai, Nomi, Ishikawa 923-1292, Japan
guoqiang@jaist.ac.jp

Abstract. We discuss secrecy of signals in signal transduction. As we have developed a basic concurrent language with interferencial coefficients, *Iπ-calculus*, to describe aberrance in biological models, a typing system for *Iπ-calculus* is proposed for achieving secrecy of signals in signal transduction. We show that this typing system guarantees that, if signal transduction typechecks, then it does not leak aberrance of signals.

1 Introduction

Signal transduction, short for ST, is the key to uncover the wild growth of cells. Aberrant ST is the cause of many diseases challenged by modern medicine, including cancers, inflammatory diseases, and so on. Formal method is one of approach to research ST. Process algebra, such as *pi* calculus and its variation, is a way to model ST system. There are several pieces of related work about modelling ST [4,5,2,3], based on *pi calculus* [1,6]. *Interference pi calculus* [8], (*Iπ- calculus*) is proposed to model aberrant ST.

When a signal mutates aberrantly, we want to know what will happen in the whole ST. We used a typing system [9], to replace the tag system which is used to label the existence of aberrance by sets computation, such as union, disjoint [8]. This typing system is simple enough to be enforced statically. It had been proved to be equivalent to the tag system in the capability of labelling the existence of aberrance [9].

In this paper, we emphasis on this typing system. An informal principle is developed for achieving secrecy of signals in ST. In particular, in the analyzing aberrance of ST, we label each protein(its domains) and each signal as either normal or aberrant. A signal-receiver can not find the signal from signal-sender is normal or aberrant during the transduction of signals. That is to say, the whole ST does not leak aberrance of signal. The notion of leaking is formalized

* The work is supported by The National Distinguished Young Scientist Fund of NNSFC (60225012), BDCC (03DZ14025), The National Nature Science Foundation of China (60473006), and The BoShiDian Research Fund (20010248033).

** The author is also supported by PPS, Paris 7 University, France.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 384–393, 2006.

in terms of testing equivalence. We point out, if ST typechecks, then it does not leak the aberrance of signals.

2 Secrecy of Signals in Signal Transduction

Signal transduction is a manner to answer the stimulation of cells outside. It is the key to uncover the wild growth of cells leading to many diseases. When the whole ST works perfectly, decisions of growth and death of cells are also made by rule and line. When some signals mutate aberrantly, however, the whole ST could be interfered. Cancers are some diseases resulting from this kind of interference of ST.

The sequel that the ST is interfered is, the growth of a cell is never controlled by growth factors outside. There exist many methods to get it, one of which is that some aberrant proteins make the cell release the growth factors into the environment. These factors can stimulate the cell which sets them free, and make it grow. Another is the aberrance of *Ras* protein. Normal *Ras* protein in the inactive state is waiting for the signal. It is activated when it receives the signal, and then sends signal to the others. After that, it could be inactivated to return the initial state. This kind of inactivity make assure that the cell just can receive finite signals.

Aberrant *Ras* protein has some difference with normal *Ras* protein. Aberrant *Ras* protein can be activated and send a signal to the others, as same as normal *Ras* protein. Aberrant *Ras* protein however can not be inactivated any more. That means, it will be always in the active state and always sends the signal to the others, even there is no real signal coming.

In a word, whatever is chosen, aberrant proteins pretend the normal proteins to transduce stimulation signals. In another word, a carrier waiting for signals does not know the category of signals.

Therefore, we have an important property about aberrance:

In signal transduction, a sender does not leak aberrance of signals.

This is also a principle in the studying of signal transduction.

3 The Pure *Iπ-calculus*

Iπ-calculus is proposed to describe aberrant ST. This section presents the pure version of *Iπ-calculus* that serves as the preliminary setting for our formal work.

We assume an infinite countable set $\mathcal{A}$ of values, an infinite countable set $\mathcal{N}$ of names and an infinite countable set $\mathcal{V}$ of variables. Let σ, ρ be functions from $\mathcal{N}$ to $\mathcal{A}$. One can think of σ as an interference function and that $\sigma(a)$ as the interference degree of a. The function ρ is a critical function and that $\rho(a)$ is the critical value of the interference degree of a. The interferential coefficient can be defined below:

Definition 1 (Interferential Coefficient). *For $a \in \mathcal{N}$, let i_a be $|\rho(a) - \sigma(a)|$. We say that i_a is the interference coefficient of a.*

Intuitively, when i_a is equal to zero, we take that a is in an aberrant state; when i_a is not zero, we think that a is still in a normal state.

Processes evolve by performing actions. In process algebra actions capabilities are introduced by prefix capabilities. In *Iπ-calculus*, we introduce two capabilities in addition to the prefix defined by *pi calculus*.

Let $a, b, \cdots$ range over the names and $x, y, \cdots$ range over the variables. We also define two symbols § and ♯ to represent the aberrance capability. Here § represents the suicide capability and ♯ the propagation capability. When a process has the suicide capability, it terminates its action immediately. And when a process has the propagation capability, it will duplicate its action infinitely.

Definition 2 (Prefix). *The prefix of Iπ-calculus are defined as follows:*

$$\pi ::= \overline{a}(b) \mid a(x) \mid \overline{a} \mid a \qquad \pi_i ::= \pi \mid [i_{\pi_i} = 0]\S(\pi_i) \mid [i_{\pi_i} = 0]\sharp(\pi_i)$$

The capability of π is the same as in *pi calculus*. $[i_{\pi_i} = 0]\S(\pi_i)$ and $[i_{\pi_i} = 0]\sharp(\pi_i)$ are the substitution capabilities. They are respectively the capabilities § and ♯ if the subject of π_i is in an aberrant state.

Definition 3 (Process). *The Iπ-calculus processes are defined as follows:*

$$P ::= 0 \mid \pi_i.P \mid \pi_i.P + \pi_i'.P' \mid P|P' \mid (\nu a)P$$

Intuitively the constructs of *Iπ-calculus* processes have the following meaning: 0 is the inert process. The prefix process $\pi_i.P$ has a single capability imposed by π_i, that is, the process P cannot proceed until that capability has been exercised. The capabilities of the sum $\pi_i.P + \pi_i'.P'$ are those of $\pi_i.P$ plus those of $\pi_i'.P'$. When a sum exercises one of its capabilities, the other is rendered void. In the composition process $P|P'$, the components P and P' can proceed independently and can interact via shared channel. In the restriction process $(\nu a)P$, the scope of the name a is restricted to P.

We write $fn(P)$ for the set of free names in process P, and $fv(P)$ for the set of free variables in P. An expression is closed if it has no free variables. Notice that a closed expression may have free names.

The reaction relation, introduced initially by Milner [1], is a concise account of computation in the pi calculus. In addition to the well-known interaction rule(Com-N), our reaction relation also includes two new rules about reactions with aberrance(Pre-§ and Pre-♯).

A *barb* is a name m, a co-name $\overline{m}$ or two primitives § and ♯. An action is a barb or the distinguished *silent action* τ. We range α, β, $\cdots$ over actions.

$$\frac{}{[i_{\pi_i} = 0]\S(\pi_i).P \xrightarrow{\S} 0}\ \text{Pre-§}\ ; \qquad \frac{}{[i_{\pi_i} = 0]\sharp(\pi_i).P \xrightarrow{\sharp} \pi_i.[i_{\pi_i} = 0]\sharp(\pi_i).P}\ \text{Pre-♯}\ ;$$

$$\frac{}{\overline{a}(b).Q \mid a(x).P \xrightarrow{\tau} Q|P\{b/x\}}\ \text{Com-N}; \qquad \frac{}{\overline{a}.Q \mid a.P \xrightarrow{\tau} Q \mid P}\ \text{Com-SN}$$

$$\frac{P \xrightarrow{\alpha} P'}{P + Q \xrightarrow{\alpha} P'}\ \text{Sum}; \qquad \frac{P \xrightarrow{\alpha} P'}{P \mid Q \xrightarrow{\alpha} P' \mid Q}\ \text{Com};$$

$$\frac{P \xrightarrow{\alpha} P' \quad a \neq \alpha}{(\nu a)P \xrightarrow{\alpha} (\nu a)P'} \text{Res;} \qquad \frac{Q \equiv P \quad P \xrightarrow{\alpha} P' \quad P' \equiv Q'}{Q \xrightarrow{\alpha} Q'} \text{Stc.}$$

The first two rules deal with reactions with aberrance: the former says that the resulting process is terminated; the latter declares that the resulting process duplicates its action infinitely. The third reaction rule deals with the interaction in which one sends a message with a channel while the other receives a message with the same channel so that they have an interactive action. Each of the reduction rules are closed in the summation, composition, restriction and structural congruence.

Next, we represent some preliminaries of testing equivalence. These notions are belong to Martín Abadi [7].

A *test* is a pair (Q, β) consisting of a closed process Q and a barb β. We say that P *passes* a test (Q, β) if and only if $(P \mid Q) \xrightarrow{\tau} Q_0 \cdots \xrightarrow{\tau} Q_n \xrightarrow{\beta} A$

For some $n \geq 0$, some processes $Q_0, \cdots Q_n$, and some process A, we obtain a testing preorder $\sqsubseteq$ and a testing euqivalence $\simeq$ on closed processes:

$P \sqsubseteq P' \stackrel{\triangle}{=} for\ any\ test\ (Q, \beta),\ if\ P\ passes\ (Q, \beta)\ then\ P' passes\ (Q, \beta)$

$p \simeq P' \stackrel{\triangle}{=} P \sqsubseteq P'\ and\ P' \sqsubseteq P$

A *strict barbed simulation* is a binary relation $\mathcal{S}$ on closed processes such that $P\mathcal{S}\mathcal{P}'$ implies:

(1) for every barb β, if $P \xrightarrow{\beta} A$ for some A, then $P' \xrightarrow{\beta} A'$ for some A',
(2) for every P_1, if $P_1 \xrightarrow{\tau} P_1$ then there exists $P_1{}'$ such that $P' \xrightarrow{\tau} P_1{}'$ and $P_1 \mathcal{S} P_1{}'$.

A *strict barbed bisimulation* is a relation $\mathcal{S}$ such that both $\mathcal{S}$ and $\mathcal{S}^{-1}$ are strict barbed simulations.

The following lemma provides a method for proving testing equivalence:

Lemma 1. *If for every closed process Q there exists a strict barbed bisimulation $\mathcal{S}$ such that $(P \mid Q)\mathcal{S}(P' \mid Q)$, then $P \backsimeq P'$*

In [7], Martín Abadi gave a simple direct proof.

4 The Typing System

This section describes rules for controlling information flow in *Iπ-calculus* calculus. Here we embody them in a typing system for *Iπ-calculus* calculus. The typing system was firstly introduced by Martin Abadi in studying security protocols [7].

In order to represent the aberrance of ST we classify signals into three classes:

- A *Normal* signal is one that takes part in the normal processes.
- An *Aberrant* signal is one that takes part in the aberrant processes.
- An *Unknown* signal could be any signal.

To simplify we define a reflexive order relation $<:$ among these three classes:

$$\mathit{Normal} <: \mathit{Unknown};\ \mathit{Aberrant} <: \mathit{Unknown}.$$

For convenience of representation, we denote M as a name or a variable. M is called *term*. Corresponding to these three classes the typed system has three kinds of assertions:

- "$\vdash \Gamma$ well formed" means that the environment Γ is well-formed.
- "$\Gamma \vdash M : T$" means that the term M is of the class T in Γ.
- "$E \vdash P : ok$" means that the process P typechecks in E.

Typing rules are given under an environment. An environment is a list of distinct names and variables with associated classifications.

Definition 4 (Typed Environment). *Typed environments are given by the following rules:*

$$\frac{}{\vdash \emptyset\ well\ formed}\mathit{Environment\ Empty}$$

$$\frac{\vdash \Gamma\ well\ formed, M \notin \Gamma}{\vdash \Gamma, M : T\ well\ formed}\mathit{Environment\ Term}$$

Having defined the environments, one can define rules for terms and processes.

Definition 5 (Terms). *The rules for terms of typing system are as follows:*

$$\frac{\Gamma \vdash M : T \quad T <: R}{\Gamma \vdash M : R}\mathit{Level\ Subsumption}$$

$$\frac{\vdash \Gamma\ well\ formed \quad M : T\ in\ \Gamma}{\Gamma \vdash M : T}\mathit{Level\ Term}$$

Intuitively the rule Level Subsumption says that a term of level *Normal* or *Aberrant* has level *Unknown* as well.

Definition 6 (Processes). *The rules for typing processes are as follows:*

$$\frac{\Gamma \vdash a : Normal \quad \Gamma \vdash b : Normal \quad \Gamma \vdash P : Ok}{\Gamma \vdash \overline{a}(b).P : Ok}\mathit{T\text{-}out}$$

$$\frac{\Gamma \vdash a : Normal \quad \Gamma \vdash x : Unknown \quad \Gamma \vdash P : Ok}{\Gamma \vdash a(x).P : Ok}\mathit{T\text{-}in}$$

$$\frac{\Gamma \vdash a : Normal \quad \Gamma \vdash P : Ok}{\Gamma \vdash \overline{a}.P : Ok}\mathit{T\text{-}sout} \quad \frac{\Gamma \vdash a : Normal \quad \Gamma \vdash P : Ok}{\Gamma \vdash a.P : Ok}\mathit{T\text{-}sin}$$

$$\frac{\Gamma \vdash a : Aberrant \quad \Gamma \vdash b : Normal \quad \Gamma \vdash P : Ok}{\Gamma \vdash \overline{a}(b).P : Ok}\mathit{T\text{-}aout}$$

$$\frac{\Gamma \vdash a : Aberrant \quad \Gamma \vdash x : Unknown \quad \Gamma \vdash P : Ok}{\Gamma \vdash a(x).P : Ok}\mathit{T\text{-}ain}$$

$$\frac{\Gamma \vdash a : Aberrant \quad \Gamma \vdash P : Ok}{\Gamma \vdash \overline{a}.P : Ok}\mathit{T\text{-}asout} \quad \frac{\Gamma \vdash a : Aberrant \quad \Gamma \vdash P : Ok}{\Gamma \vdash a.P : Ok}\mathit{T\text{-}asin}$$

$$\frac{\Gamma \vdash a : Aberrant \quad \Gamma \vdash b : Normal \quad \Gamma \vdash P : Ok}{\Gamma \vdash [i_a = 0]\S(\overline{a}(b)).P : Ok}\,T\text{-}kout$$

$$\frac{\Gamma \vdash a : Aberrant \quad \Gamma \vdash x : Unknown \quad \Gamma \vdash P : Ok}{\Gamma \vdash [i_a = 0]\S(a(x)).P : Ok}\,T\text{-}kin$$

$$\frac{\Gamma \vdash a : Aberrant \quad \Gamma \vdash P : Ok}{\Gamma \vdash [i_a = 0]\S(\overline{a}).P : Ok}\,T\text{-}ksout \quad \frac{\Gamma \vdash a : Aberrant \quad \Gamma \vdash P : Ok}{\Gamma \vdash [i_a = 0]\S(a).P : Ok}\,T\text{-}ksin$$

$$\frac{\Gamma \vdash a : Aberrant \quad \Gamma \vdash b : Normal \quad \Gamma \vdash P : Ok}{\Gamma \vdash [i_a = 0]\sharp(\overline{a}(b)).P : Ok}\,T\text{-}pout$$

$$\frac{\Gamma \vdash a : Aberrant \quad \Gamma \vdash x : Unknown \quad \Gamma \vdash P : Ok}{\Gamma \vdash [i_a = 0]\sharp(a(x)).P : Ok}\,T\text{-}pin$$

$$\frac{\Gamma \vdash a : Aberrant \quad \Gamma \vdash P : Ok}{\Gamma \vdash [i_a = 0]\sharp(\overline{a}).P : Ok}\,T\text{-}psout \quad \frac{\Gamma \vdash a : Aberrant \quad \Gamma \vdash P : Ok}{\Gamma \vdash [i_a = 0]\sharp(a).P : Ok}\,T\text{-}psin$$

$$\frac{\vdash \Gamma \ well\ formed}{\Gamma \vdash 0 : Ok}\,T\text{-}nil \quad \frac{\Gamma, a : Normal \vdash P : Ok, \quad a \notin dom(\Gamma)}{\Gamma \vdash (\nu a)P : Ok}\,T\text{-}res$$

$$\frac{\Gamma \vdash P : Ok \quad \Gamma \vdash Q : Ok}{\Gamma \vdash P \mid Q : Ok}\,T\text{-}com \quad \frac{\Gamma \vdash P : Ok \quad \Gamma \vdash Q : Ok}{\Gamma \vdash P + Q : Ok}\,T\text{-}sum$$

$$\frac{\Gamma \vdash a : Unknown\ \Gamma \vdash b : Noraml\ \Gamma \vdash P : Ok}{\Gamma \vdash P\{b/x\} : Ok}\,T\text{-}Sub \qquad \frac{\Gamma \vdash P : Ok\ Q \equiv P}{\Gamma \vdash Q : Ok}\,T\text{-}stc$$

5 Secrecy of Signals by Typing

As mentioned in the section 2, an important principle in the modelling signal transduction is to guarantee that category of signals can not be detected. In this section, we show that a signal-receiver can not distinguish the difference of signal using our typing system. The original idea is from [7]. In that paper, Martín Abadi applies the similar typing system to *Spi calculus*, and uses it to analyzes security protocols. Our main result says that if only variables of level $Unknown$ and only names of level $Normal$ are in the domain of the environment E, if σ and σ' are two substitutions of values for the variables in E, and if P typechecks, then $P\sigma$ and $P\sigma'$ are testing equivalence.

We write $E \vdash \sigma$ when $\sigma(x)$ is a closed term such that $fn(\sigma(x)) \subseteq dom(E)$ for every $x \in dom(E)$.

Lemma 2. *Suppose that E is an environment which all variables in $dom(E)$ are of level $Unknown$, $E \vdash P : ok$ and $E \vdash \sigma$. Then we have*

(1) *if $P\sigma \xrightarrow{\tau} Q'$, then there exists a process Q such that*
- $Q' = Q\sigma$
- $E \vdash Q : ok$
- $P\sigma' \xrightarrow{\tau} Q\sigma'$ *whenever* $E \vdash \sigma'$

(2) *if $P\sigma \xrightarrow{\beta} A'$, there exists a process A such that*
- $A' = A\sigma$
- $E \vdash A : ok$
- $P\sigma' \xrightarrow{\beta} A\sigma'$ *whenever* $E \vdash \sigma'$

Theorem 1. *Given an environment E, suppose that all the variables in $dom(E)$ are of level Unknown. Suppose further that $E \vdash \sigma$ and $E \vdash \sigma'$. Then the relation*

$$\{(P\sigma, P\sigma') \mid E \vdash P : ok\}$$

is a strict barbed bisimulation.

This conclusion means that a signal-receiver can not distinguish $P\sigma$ and $P\sigma'$, so it can not detect the difference of signals.

6 An Example in Signal Transduction

In order to illustrate the use of our typing rules, we consider as an example of aberrance of *Ras* protein. *Ras* protein is an important protein in the well-studied *RTK-MAPK* pathway.

Fig.1 gives an example of *Ras* Activation of the ST pathway, *RTK-MAPK*. At the normal state, the protein-to-protein interactions bring the *SOS* protein close to the membrane, where *Ras* can be activated. *SOS* activates *Ras* by exchanging *Ras*'s *GDP* with *GTP*. Active *Ras* interacts with the first kinase in the *MAPK* cascade, *Raf*. *GAP* inactivates it by the reverse reaction. When *Ras* mutates aberrantly, it does not have any effect on the *Ras*'s binding with *GTP* but will reduce the activity of the *GTP* hydrolase of *Ras* and lower its hydrolysis of *GTP* greatly; in the meantime *Ras* will be kept in an active state; it keeps activating the molecule, inducing the continual effect of signal transduction, which result in cell proliferation and tumor malignancy. Aviv Regev and his colleagues have given the representation of normal *RTK-MAPK* using the *pi* calculus [4]. We had given the representation of the aberrant *Ras* protein using *$I\pi$-calculus*.[8].

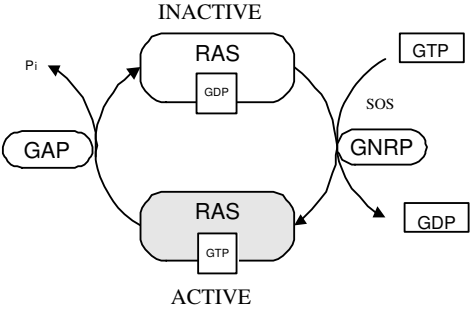


Fig. 1. *Ras* Activation

The interpretation of *Ras* in the *$I\pi$-calculus* can be done in the following manner: The system defined in (1) is a collection of concurrently operating molecules, seen as processes with potential behavior. Here the operator | is the concurrent

combinator. In biological models, all the names are bounded, for simplicity, we here only list the aberrant names for restriction :

$$SYSTEM ::= (\nu s_{ACTSWI\_I})(\nu bbone_{ACTSWI\_I})(\nu sg_{ACTSWI\_II}) \\ (RAS \mid SOS \mid GAP \mid RAF) \quad (1)$$

A protein molecule is composed of several domains, each of which is modelled as a process as well. In (2) through (5) the detailed $I\pi$-*calculus* programs for the proteins *Ras*, *SOS*, *Raf* and *GAP* are given:

$$RAS ::= INASWI\_I \mid INASWI\_II \quad (2)$$
$$SOS ::= S\_SH3\_BS \mid S\_GNEF \quad (3)$$
$$RAF ::= R\_Nt \mid R\_ACT\_BS \mid R\_M\_BS \\ \mid INA\_R\_Ct \mid R\_ATP\_BS \quad (4)$$
$$GAP ::= sg(x).\overline{x}(gdp).GAP \quad (5)$$

The molecules (or domains) interact with each other based on their structural and chemical complementarity. Interaction is accomplished by the motifs and residues that constitute a domain. These are viewed as channels or communication ports of the molecule:

$$INASWI\_I ::= \overline{bbone}.ACTSWI\_I \quad (6)$$
$$INASWI\_II ::= \overline{sg}(rs\_1).rs\_1(y).bbone.ACTSWI\_II \quad (7)$$
$$S\_GNEF ::= bbone.S\_GNEF + sg(z).\overline{z}(gtp).S\_GNEF \quad (8)$$
$$S\_SH3\_BS ::= \overline{bbone}.S\_SH3\_BS \quad (9)$$

The following interactions are possible:

$$INASWI\_I \mid S\_GNEF \longrightarrow ACTSWI\_I \mid S\_GNEF \quad (10)$$
$$INASWI\_II \mid S\_GNEF \longrightarrow^{*} bbone.ACTSWI\_II \mid S\_GNEF \quad (11)$$
$$bbone.ACTSWI\_II \mid S\_SH3\_BS \longrightarrow ACTSWI\_II \mid S\_SH3\_BS \quad (12)$$

The interaction (10) shows that the domain $INASWI\_I$ of *Ras* is activated by the domain of S_GNEF of *SOS*. The interaction (11) and (12) show that the domain $INASWI\_II$ of *Ras* is activated by the domain S_GNEF and S_SH3_BS of *SOS*.

The detailed $I\pi$-*calculus* programs for activated domains, $ACTSWI\_I$ (Aberrant), $ACTSWI\_II$ (Aberrant) of the protein *Ras* and the domain R_Nt of *Raf* are defined in (13) through (15):

$$ACTSWI\_I^{*} ::= (\nu s)[i_s = 0]\sharp(\overline{s}(rs\_2).\overline{rs\_2}).INACTSWI\_I + \\ (\nu bbone)[i_{bbone} = 0]\S(\overline{bbone}).INASWI\_I \quad (13)$$
$$ACTSWI\_II^{*} ::= (\nu sg)[i_{sg} = 0]\S(\overline{sg}(r\_swi\_1)).r\_swi\_1(h).\overline{bbone}. \\ INACTSWI\_II \quad (14)$$
$$R\_Nt ::= s(i).i.ACTR\_Nt \quad (15)$$

The processes so defined have the following interactions:

$$ACTSWI\_I^* \mid R\_Nt \longrightarrow^* (\nu s)[i_s = 0]\sharp(\overline{s}(rs\_2).\overline{rs\_2}).INACTSWI\_I \mid ACTR\_Nt \quad (16)$$

$$ACTSWI\_II^* \longrightarrow 0 \quad (17)$$

The interaction (16) shows that the active domain $ACTSWI\_I$ of *Ras* interacts with the domain R_Nt of *Raf*, but it can not be inactivated any more. (17) shows that the domain $ACTSWI\_II$ of *Ras* can not inactivated by *GAP*.

In order to indicate how the process $SYSTEM$ typecheks, we annotate its bound names and variables with there levels, as they are introduced.

Let E be an environment, where the variables are $Unknown$ levels:

$$x : Unknown,\ z : Unknown, h : Unknown\ i : Unknown.$$

The bound names with their $Normal$ levels:

$$sg_{GAP} : Normal,\ bbone_{INASWI\_I} : Normal,\ sg_{INASWI\_II} : Normal,$$
$$rs\_1_{INASWI\_II} : Normal, bbone_{INASWI\_II} : Normal, bbone_{S\_GNEF} : Normal,$$
$$sg_{S\_GNEF} : Normal,\ bbone_{S\_SH3\_BS} : Normal,\ rs\_2_{ASWI\_I} : Normal,$$
$$r\_swi\_1_{ACTSWI\_II} : Normal,\ bbone_{ASWI\_II} : Normal,\ s_{R\_Nt} : Normal.$$

The bound names with there *Aberrant* levels:

$$s_{ACTSWI\_I} : Aberrant, bbone_{ACTSWI\_I} : Aberrant, sg_{ACTSWI\_II} : Aberrant.$$

We define:

$$Ras \triangleq (\overline{bbone} : Normal)[(\overline{s} : Aberrant)(rs\_2).(\overline{rs\_2} : Normal).INACTSWI\_I + (\overline{bbone} : Aberrant).INASWI\_I] \mid (\overline{sg} : Normal)(rs\_1).(rs\_1 : Normal)(y).(bbone : Normal).(\overline{sg} : Aberrant)(r\_swi\_1).(r\_sw\_1 : Normal)(h).(\overline{bbone} : Normal).INACTSWI\_II$$

$$SOS \triangleq (\overline{bbone} : Normal).S\_SH3\_BS \mid [(bbone : Normal).S\_GNEF + (sg : Normal)(z).(z : Unknown)(gtp).S\_GNEF]$$

$$RAF \triangleq (s : Normal)(i).(i : Unknown).ACTR\_Nt \mid ...$$

$$GAP \triangleq (sg : Normal)(x).(\overline{x} : Unknown)(gdp).GAP$$

Finally, in the given environment E, we set

$$SYSTEM \triangleq (\nu s_{ACTSWI\_I} : Aberrant)(\nu bbone_{ACTSWI\_I} : Aberrant)(\nu sg_{ACTSWI\_II} : Aberrant)(RAS \mid SOS \mid GAP \mid RAF) \quad (18)$$

It is easy to find $E \vdash SYSTEM : Ok$. By Theorem 1, as a consequence of the typechecking, we obtain that $SYSTEM$ does not reveal the aberrance of *Ras* protein.

7 Future Prospects

The typing system we introduced is very simple but strong. It can applied not only into analyzing security protocols but also in the study of signal transduction with exception, which is also opening up new possibilities in modelling of biochemical systems.

To make quantitative analysis is another important step for studying biochemical systems. So far, our *Iπ-calculus* is concerned about qualitative analysis. Could it do some quantitative analysis? Two functions ρ and σ which are used to describe some quantitative properties of proteins will be found its value in future work.

References

1. Milner, R., Parrow, J., Walker, D.: A Calculus of Mobile Processes, parts I and II. In: Information and Computation. (1992)1-77
2. Priami, C., Regev, A., Silverman, W., and Shapiro,E.: Application of a stochastic name passing calculus to representation and simulation of molecular processes. In: Information Processing Letters. **80**(2001)25-31
3. Regev, A.: Representation and simulation of molecular pathways in the stochastic pi calculus. In: Proceedings of the 2nd workshop on Computation of Biochemical Pathways and Genetic Networks. (2001)
4. Regev, A., Silverman, W., and Shapiro, E.: Representing biomolecular processes with computer process algebra: pi calculus programs of signal transduction pathways. In: http://www.wisdom.weizmann.ac.il/ aviv/papers.htm (2000)
5. Regev, A., Silverman, W., and Shapiro, E.: Representation and simulation of biochemical processes using the pi calculus process algebra. In: Proceedings of the Pacific Symposium of Biocomputing. **6**(2001)459-470
6. Sangiorgi, D., and Walker, D.: The pi calculus: a Theory of Mobile Process. In: Cambridge University Press. (2001)
7. Abadi, M.: Secrecy by Typing in Security Protocals. In Proceedings of Theoretical Aspects of Computer Software, Third International Symposioum. **1281**(1997)611-638
8. Zhang, M., Li, G., and Fu, Y., et al.: Representation of the Signal Transduction with Aberrance Using Ipi Calculus. In : Computational and Information Science: First International Symposium, LNCS**3314**(2004)477-484
9. Zhang, M., Li, G, and Fu, Y., et al.: Typing Aberrance in Signal Transduction. In: ICNC 2005, LNCS **3612**(2005)668-677

The Coarse-Grained Computing P2P Algorithm Based on SPKI

Yong Ma and Yumin Tian

School of Computer Science & Technology, Xidian University,
Xian, 710071, China
ymtian@mail.xidian.edu.cn,
mywuda@126.com

Abstract. Distributed components, such as CORBA and COM+, implement the two-way invocation between the client and server using RPC in the distributed network. RPC requires that server must be certain and online all the time, which will endow the server with heavy burden and the system with single dependence on the server. The task certification and result certification, based on SPKI, are proposed to transform the RPC to Request/Response of XML-formed Certification. DHT is used to store and look up the certificates, which enable MR2CG (the Method of Coarse-Grained P2P Computing Based on XML Request/Response) to get rid of the limitation of the server and support the different kinds of Coarse-grained P2P computing.

1 Introduction

The central server can manage all the clients to collaboratively work in the traditional client/server or browser/server net. Though it is easy to implement the above structures, what cannot be neglected are the busy server and one point failure of the server resulting in the malfunction of the system. The server in the Client/Server-based coarse-grained computing must allot the subtasks and cope with these results of the subtasks. Consequently, all the clients must know the server address to set up the communication in advance and the server must be always on line. All these characteristics determine the Client/Server net can not well match the P2P net featuring dynamism and symmetry.

SPKI/SDSI has resolved naming and authorization in the distributed network and presented one relatively comprehensive trust management. And DHT describes how to look up the resource in an efficient model in the distributed network. These lay a solid foundation for managing the XML-formed messages efficiently. The existing RPC is one special kind of message, involving some necessary parameters, which informs the remote object implementing a task [1]. SPKI-based CGP2PC (Coarse-Grained P2P Computing) has been presented in this paper to meet the following goals:

- Constructing the system, in which the server is eliminated and all the peers act as both server and client to meet the dynamism and symmetry of the P2P net,
- Guarantying the security of the task and result,
- Auditing the results.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 394 – 401, 2006.

2 Related Works

Distributed computing can not only be put into applications, but also boom in science research on multi-agent fields [2, 3]. Distributed computing is based on communication protocols (TCP, UDP) and middlewares (DCOM, CORBA). TCP and UDP require that communication components must know the counter's IP and port. And their reference implementations offer the different invocation interfaces of the client and server. So TCP and UDP are implemented on the basis of the Client/Server structure. The CORBA of OMG and DCOM of Microsoft are standard distributed middleware. The client request and server response are carried out via ORB in CORBA [4], and mutual invocations between the server and the client component are executed via the connect point in DCOM [5]. Both CORBA and DCOM are able to implement the binary interoperation, offer the neutral interface, and provide transparency of the location. However, they are based on Client/Server structure and do not suit well the P2P network.

There are many exampled applications based on Client/Server structure. Seti@Home[6] is one of the SETI (Search for Extraterrestrial Intelligence) projects. Other examples include solving the RSA cryptographic challenge on idle cycles of personal computers and finding large Mersenne prime numbers. The Verbeke of SUN Company has presented one framework for peer-to-peer distributed computing in a heterogeneous and decentralized environment, in which peers can be divided into three kinds of roles: TRM, TD, and TW. However, TDs take the server role in charge of allotting the subtasks and managing all the TWs [7]. However, TWs, implementing special computing task, will only exchange the messages with the peer performing the task assignment, which is not able to suit well the characteristics of P2P network and limits the computing efficiency.

SPKI/SDSI presents one method on how to construct the authentication and authorization certificates [8]. It will take a long time for one peer to search the resource in the distributed and decentralized network because resources and replications are stored in diverse peers. DHT (Distributed Hash Table) [9] is one high efficient method to search the resources in the distributed network, and ConChord [10] and CAN [11] are the typical cases of DHT.

3 MR2CG

3.1 Definitions

Let the task be S and the result be R. The task S can be divided into N subtasks: St(1)…St(N). Let the corresponding results of N subtasks be Rt(1) … Rt(N), run time be T(1) … T(N), and the parameters be Pt(1) … Pt(N). There will be

- Parameters independency: $Pt(1)\cap\ldots\cap Pt(N)=NULL$
- Order independency: If the task is implemented in different orders, the result makes no difference.
- Result independency: $Rt(1)\cap\ldots\cap Rt(N)=NULL$
- Execution independency: Any single subtask can be carried out independently.

CGC (Coarse-Grained Computing): the task and subtasks St(1)…St(N) can meet the above four kinds of independency.

CICGC (Completely Independent Coarse-Grained P2P Computing): the result of the task is equal to the result of a subtask, that is, it can be defined in the mathematics form: $\exists(m)\{R\equiv Rt(m), 1\leq m\leq N\}$. Brute decryption DES key and Eight Queens are the typical examples.

PCCGC (Partly Collaborative Coarse-Grained P2P Computing PCCGC) The result of the task is composed of all the subtasks' results, that is, it can be defined in the mathematics form: $\{R=\cup Rt(i), 1\leq i\leq N\}$. Sorting the large scale data and finding the prime numbers are the standing examples.

SPKI name certificate can be used to express task S and subtasks St(1)…St(N) binding task and issuer together. To describe the results Rt(1)…Rt(N) binding result and executor together. So task and result certificates are:

- Task certificate <Issuer, Task, TID, PTID, Validity>: Issuer is a public key or hash value of the Issuer, TID is the Task ID, PTID is the parent task ID of TID, and Validity is the period of certificate's validity.
- Result certificate <Executor, Result, TID, PTID, Validity>: Executor is a public key or hash value of the Executor, Result is the result of the TID task, TID is the Task ID, PTID is the parent task ID of TID, and Validity is the period of the certificate's validity.

3.2 MR2CG Structure

MR2CG, depicted as Fig.1, comprises three layers from top to bottom: application layer, DHT layer and base layer. Application layer can mainly arrive at producing the task and result certificate, and implementing the computing. DHT layer can be in charge of storing and locating all the certificates, reconstructing the DHT ring when the peer joins and quits the group, and providing the interface of partitioning task, allotting task, executing task and expiring the certificate. The base layer is composed of net, software and hardware and offers the basic and necessary services.

DHT layer revises the find_successor(id), find_predecessor(id) and Closet_ preceding_finger(id) to find the task certificate. Afterwards, DHT layer will return the task certificate from peer storing the certificate to the request peer, and pick the task certificate in the order of own-first and succeed-second. As a result, DHT layer can implement the interface of "allot".

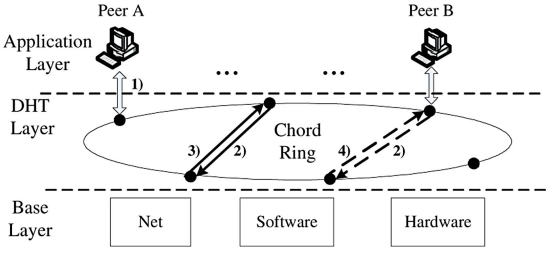


Fig. 1. MR2CG structure, 1) partition, 2) allot, 3) execute, and 4) expire

Every peer in MR2CG has three sets: FTS (Free Task Set), ATS (Allotted Task Set) and CTS (Completed Task Set). The process of MR2CG can be divided into three parts: partition task, allot task and execute task. The corresponding operations in the three periods are Partition, Allot and Execute.

Partition: Creator partitions the task T and plugs the TaskCert(No) into the FTS in the peer StorePeer.

$$\begin{gathered}\text{Partition} = \text{Creator} \xrightarrow{\text{TaskCert(No)}} \text{Store Peer} \\ \text{Store Peer} : \text{FTS} = \text{FTS} \cup \text{TaskCert(No)}\end{gathered} \tag{1}$$

Allot: Request Peer applies for the new free task certificate and Allot Peer will dispatch TaskCert(No) to the Request Peer. Allot Peer will change the state of the TaskCert(No) from FTS to CTS.

$$\begin{gathered}\text{Allot} = \text{Allot Peer} \xrightarrow{\text{TaskCert(No)}} \text{Request Peer} \\ \text{Allot Peer: FTS} = \text{FTS- TaskCert(No)}, \ \text{ATS} = \text{ATS} \cup \text{TaskCert(No)}\end{gathered} \tag{2}$$

Execute: If Execute Peer finishes the TaskCert(No), it will produce the result certificate and insert the result certificate into the CTS in the Allot Peer. And then the Allot Peer change the state of the TaskCert(No) from ATS to CTS.

$$\begin{gathered}\text{Execute} = \text{Execute Peer} \xrightarrow{\text{ResultCert(No)}} \text{Allot Peer} \\ \text{Allot Peer:ATS=ATS-TaskCert(No),Allot Peer:CTS=CTS} \cup \text{ResultCert(No)}\end{gathered} \tag{3}$$

Expire: If Allot Peer does not receive the result certificate corresponding to TaskCert (No) during the Validity, it will change the state of the TaskCert(No) from ATS to FTS.

$$\begin{gathered}\text{Expire Allot Peer} \xrightarrow{\text{TaskCert(No)}} \text{Allot Peer} \\ \text{Allot Peer: ATS} = \text{ATS- TaskCert(No)}, \ \text{FTS} = \text{FTS} \cup \text{TaskCert(No)}\end{gathered} \tag{4}$$

4 Interface Design

DHT layer in MR2CG mainly provides four function invocations: Partition, Allot, and Execute. One implementation algorithm based on Chord ring and one sum case will be given in the following part.

4.1 Partition Invocation

Partition invocation aims at splitting the task into the smaller grained subtasks, generating task certificates corresponding to the subtasks, and inserting these task certificates into the peers in the chord ring. The specialized steps can be listed as follows:

- According to the one way hash function, all the online peers are mapped into the Chord space.
- The peer setting up the computing task will divide the task into subtasks, which will be mapped into the Chord space according to the ID of the task certificate generated by the same direct hash function.

– These task certificates are inserted into the Chord ring and put into the FTS of the correspondent peer.
– If the peer wants to further divide these subtasks into smaller subtasks, the step 2 and 3 are going to be implemented. Or else, the tag PTID in the task certificate will be same as the TID.

Chord instance that sum form 1 to 1000 has been partitioned is illustrated in Fig.2. Chord data structure has been modified and the state item is added up to the finger table to express what the current state of the task certificate is. How to generate the finger table has been explicitly given in [13]. The online peers store all the task certificates whose states are FTS indicating that all the tasks are free.

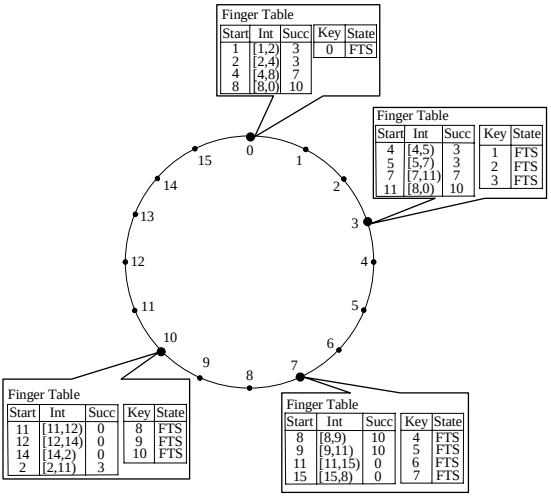


Fig. 2. Chord ring after implementing the task partition

From the above analysis, the partition interface holds the following characters:

– supporting the tree-like hierarchy task partition,
– every peer can determine the grain size and the number of the subtasks partitioned,
– the different level of task certificates uniformly managed by Chord,
– Task partition is also one coarse-grained computing and one parallel algorithm, whose time complexity is determined by the depth of the task partition tree and not influenced by the number of the subtasks.

4.2 Allotting Task

Chord only offers one lookup service so that one peer can search the value according to the key. If one peer apply for one free subtask, MR2CG will only need dispatch one arbitrary free subtask instead of one special free subtask. We modify the Chord algorithm to meet this requirement.

Theorem 1: One peer firstly accesses the succeed peer in the first row of its finger table, then access the succeed peer in the first row of its succeed peer's finger table. In this order, the peer can visit all the active peers as well as all the task certificates stored in all the online peers.

Proof: If all the online peers are sorted according to the Chord space order as Peer(0), Peer (3),…, Peer(i), Peer(j), …, Peer(M). So all these peers can construct one Chord ring: Peer(0), Peer (3),…, Peer(i), Peer(j), …, Peer(M), Peer(0). Let us say, one peer be peer(i), the next online peer be peer(j) in the Chord space. The first row of peer(i)'s finger table is $[i+2^0,i+2^1] \bmod 2^m$ and the peer(j) meets $j>=(i+2^0=i+1) \bmod 2^m$. At the same time, Peer(j) is the minimal peer among the Peer(j), …, Peer(M) that are the succeed peers of the Peer(i), that is, Peer(j) is the direct succeed peer of the peer(i). So the Peer(j) is the succeed peer in the first row of Peer(i)'s finger table. All the online peers can be visited by accessing the succeed peer in the first row of the finger table in the order described in Theorem 1. □

On the basis of Chord, the "Allot_FreeTask" function dispatching the free atom tasks is added to Chord algorithm. How to search the free atom task is depicted in Fig. 3 as follows:

- The requestor will implement the free atom subtasks stored in the local peer;
- The requestor will search the succeed peers to request one free atom task in accord with Theorem 1, and the responder will change the state of the allotted atom task to ATS;
- The responder will send the free atom task to the requestor.

```
get the free task certificate form Chord ring by the peer n.
   TaskCert Allot_FreeTask(n)
       if((q = HasFTS(n)>-1) return GetFTS(n, q);
       m = n;
       while(HasFTS(m)==-1 and m <> n)
           m = m.finger[0].successor;
        if(HasFTS(n)>-1) return GetFTS(m)
        else return null;
check the peer id whether to have the free atom task
  int HasFTS(id)
       for( i=0; i<id.FTS.Size; i=i+1)
           if(id.FTS[i].TID <> id.FTS[i].PTID) return i;
       return -1;
   get the no i free atom task from the peer id
   TaskCert GetFTS(id,i)
       AC = id.FTS[i]
       id.ATS = id.ATS + AC
       id.FTS = id.FTS - AC;
       return AC
```

Fig. 3. The algorithm of obtaining the free atom task from Chord rings by peer n

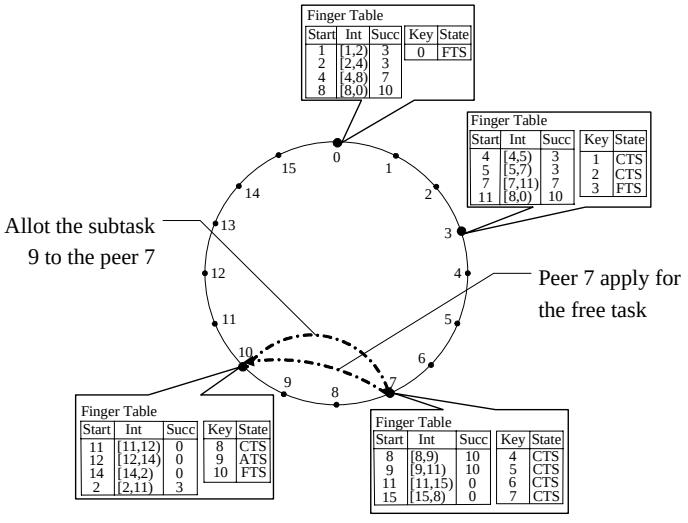


Fig. 4. Requesting and allotting free task in sum instance

Requesting and allotting one task are depicted in Fig. 4. The entire task 4, 5 and 6 in the finger table of the peer 7 have been implemented and their states are CTS. Peer 7 applies for one free atom task to its succeed peer 10, then peer 10 returns one free atom task 9, whose state is changed from FTS to ATS, to the request peer 7.

4.3 Implementing Task

Execute interface mainly generates the result certificate according to the task certificate and inserts the result certificate into the peer which stores the correspondent task certificate. Execute interface will perform the following steps:

- The result certificate will be generated according to the task certificate. If the task certificate is standing for one atom task, the result certificate will be directly inserted into the peer who stores the correspondent task certificate, and the peer will change the state of the task certificate into CTS.
- The peer storing the task certificate will inform the peer storing the parent certificate that this subtask has been completed. The peer storing the parent certificate can be searched by the peer storing the task certificate through the lookup service according to the tag PTID.
- If all the subtasks are completed, the peer storing the parent task certificate will generate the result certificate in accord with the parent task certificate and Step 1 and Step 2 will be executed.

The Execute interfaces of CICGP and PCCGP have different reference implementations. If one peer sets up one CICGP task, the peer will inform all the peers to wrap up the task upon receiving and verifying one result certificate that claims the result is successful. If the work is PCCGP, the peer can notify all the peers to finish the task until he receives all the result certificates.

5 Conclusions

MR2CG, through task certificate and result certificate, allots the tasks and submits the results for P2P distributed network, which can eliminate the limitations of the server and gear to the dynamism of P2P network. MR2CG can locate the task and result certificates to dispatch and manage the subtasks replacing the online server. All the peers can carry out the computing to improve the efficiency in MR2CG. SPKI certificates are tailored into MR2CG to perform the trust management and identity audit. MR2CG can make the coarse-grained P2P computing in a much more efficient manner than RPC and Verbeke.

References

[1] Sun Microsystems. : Remote Procedure Call Protocol Specification RFC 1057, (1988)
[2] Liu, J., Zhong,W., and Jiao, L.: A multiagent evolutionary algorithm for constraint satisfaction problems. IEEE Trans. Syst., Man, and Cybern. B, vol. 36, no. 1, (2006). 54-73
[3] Zhong, W., Liu, J., Xue, M., and Jiao, L.: A multiagent genetic algorithm for global numerical optimization. IEEE Trans. Syst., Man, and Cybern. B,vol. 34, no 2,(2004). 1128-1141
[4] OMG. CORBAservices: Common Object Services Specification. (1999)
[5] MICROSOFT. DCOM: Distributed Component Object Model Protocol – DCOM/1.0. (1996)
[6] Sullivan, W.T., Werthimer, D., Bowyer, S., Cobb, J., Gedye, D., and Anderson, D.: A new major SETI project based on Project Serendip data and 100,000 personal computers, in Astronomical and Biochemical Origins and the Search for Life in the Universe. Proc. of the Fifth Intl. Conf. on Bioastronomy. (1997)
[7] Verbeke, J., Nadgir, N., Ruetsch, G., Sharapov, I. and Parashar, M.: Framework for Peer-to-Peer Distributed Computing in a Heterogeneous. Decentralized Environment GRID, LNCS 2536, (2002). 1-12
[8] Ellison, C., Frantz, B., Lampson, B., Rivest, R., Thomas, B., and Ylonen, T.: SPKI Certificate Theory, RFC 2693. (1999)
[9] Plaxton, C., Rajaraman, R., and Richa, A.: Accessing nearby copies of replicated objects in a distributed environment. In Proceedings of the ACM SPAA. (1997)
[10] Sammer, A., Dwaine, E.C., Chuang-Hue, M., and Steven, R.: ConChord: Cooperative SDSI Certificate Storage and Name Resolution. In First International Workshop on Peer-to-Peer Systems (IPTPS). (2002). 141-154
[11] Ratnasamy, S., Francis, P., Handley, M., Karp, R., and Shenker, S.: A Scalable Content-Addressable Network. In ACM SIGCOMM. (2001)

Clonal Selection Detection Algorithm for the V-BLAST System

Caihong Mu and Mingming Zhu

Institute of Intelligent Information Processing, PO Box 224,
Xidian University, Xi'an, 710071, China
caihongm@mail.xidian.edu.cn

Abstract. A novel detection algorithm for V-BLAST system is proposed, based on the clonal selection theory and the idea of immune evolution. The complexity of the clonal selection detection algorithm for V-BLAST system (CA-VBLAST) is analyzed and the bit error ratio (BER) performance is verified via computer simulations. Simulation results show that the BER performance of CA-VBLAST detector with proper algorithm parameters is comparable with the maximum likelihood (ML) detector which presents the best BER performance but the highest computational complexity, and our CA-VBLAST detector obtains a much more decreased complexity.

1 Introduction

In the past few years, information theory has shown that multiple-input multiple-output (MIMO) systems can provide enormous capacities, provided that the multipath scattering of wireless channels is exploited with appropriate space-time processing [1-3]. The diagonally-layered space-time architecture proposed by Foschini [4], now known as diagonal BLAST (Bell Laboratories layered space-time) or D-BLAST, is one such approach. D-BLAST utilizes multi-element antenna arrays at both transmitter and receiver and an elegant diagonally-layered coding structure in which code blocks are dispersed across diagonals in space-time. This processing structure leads to theoretical rates which grow linearly with the smaller number of transmit antennas and receive antennas, and these rates can approach 90% of Shannon capacity. However, this system requires a complex coding structure that makes the detection procedure complicated. Then Vertical BLAST (V-BLAST) [5] has been proposed as a simplified version of D-BLAST, which has been implemented in real time in the laboratory. Bell labs have demonstrated spectral efficiencies of 20-40 bps/Hz at average SNRs ranging from 24 to 34 dB. Although these results were obtained in a relatively benign indoor environment, that spectral efficiencies of this magnitude are unprecedented and simply unattainable using traditional techniques. For these causes, we take V-BLAST into account.

Many V-BLAST detectors have been proposed in recent years and some classical detection algorithms are introduced in [6]. The classical Golden detector was proposed by Golden [7], but its computational complexity is high for existing many times of pseudo inverse and sort operations; the MMSE-OSIC detection algorithm [6]

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 402 – 411, 2006.

which uses MMSE nulling criterion has a good BER in the V-BLAST system with $M = N$ (M :the number of transmit antennas , N : the number of receive antennas), but its BER will worsen if $M < N$ compared with Golden algorithm; the detector combining interference cancellation with ML, which achieved a tradeoff between performance and computational complexity, was given by Ashish Bhargave [8], but a low transfer rate was resulted from the requirement of a number of training sequences for the estimation of channel matrix.

In this paper, a new detector based on clonal selection algorithm for V-BLAST system（CA-VBLAST）is presented, which obtains a comparable BER performance and a much more decreased complexity compared with ML detector.

This paper is organized as follows: Section 2 describes the system model. In Section 3, we introduce the clonal selection detection algorithm for V-BLAST system. Section 4 gives the complexity analysis. Section 5 shows the simulation results and performance analysis. Concluding remarks are given in Section 6.

2 System Model

We consider a V-BLAST system equipped with M transmit antennas and N receive antennas. The high-level block diagram of this system is shown in Figure.1. A single data stream is demultiplexed into M substreams, and each stream is then encoded into symbols and fed to its respective transmitter. Transmitter $1-M$ operate co-channel at symbol rate of $1/T$ symbols/s, with synchronized symbol timing. Regardless of the number of transmit antennas, the total radiated power is constant and the power launched by each transmitter is proportional to $1/M$. Receivers $1-N$ operate individually and co-channel, each receiving the signals radiated from all M transmit antennas. We assume that transmissions are organized into bursts of L symbols. We take the quasi-stationary viewpoint that the channel time variation is negligible over

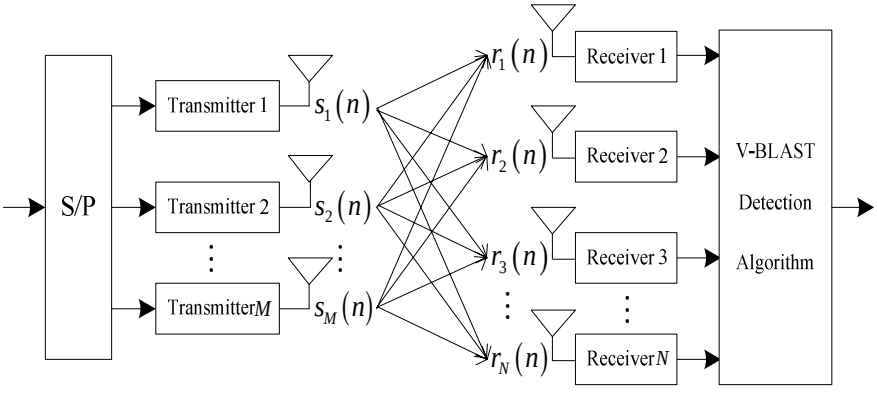


Fig. 1. System model of V-BLAST

the L symbol periods comprising a burst, and the channel is estimated accurately. For simplicity, flat fading is assumed. We also assume symbol-synchronous receiver sampling and ideal timing.

In flat fading channel, the received signal can be expressed as

$$r(n) = Hs(n) + n(n). \tag{1}$$

where $s(n) = [s_1(n), s_2(n), \cdots, s_M(n)]^T$ denote the vector of transmit symbols; $r(n) = [r_1(n), r_2(n), \cdots, r_N(n)]^T$ denote the corresponding received signals; $H = (h_{ij})_{N \times M}$ is the matrix channel transfer function, where $h_{ij}, i = 1, 2 \cdots, N$, $j = 1, 2, \cdots, M$ is the (complex) transfer function from transmitter j to receiver i, and $M \le N$. $n(n)$ is the additional Gaussian white noise vector (AWGN) with components drawn from IID wide-sense stationary processes with variance σ^2, and $E(nn^H) = N_0 U$, U is the identity matrix.

3 Clonal Selection Algorithm Based V-BLAST Detection

3.1 Clonal Selection Theory

Clonal selection theory was proposed by Burnet in 1959, and since then different clonal selection algorithms were proposed one after the other [9-14]. These human immune systems (HIS) realize researches which are uncorrelated to themselves by encoding and provide immune mechanisms such as noise endurance, unsupervised study, self-organization, memorizing and so on. HIS has been widely used in such areas as Anomaly Detection, Pattern Recognition, Job shop Scheduling, Optimization or Engineering Optimization and so on for the better ability of problem solving. Licheng Jiao, Haifeng Du and Maoguo Gong applied HIS to detection of communication systems and proposed immune clonal selection algorithm for multiuser detection in DS-CDMA communication systems [15]. It obtained a good detection performance with a low computational complexity. Inspired by the ideas above, we apply the clonal selection algorithm to the detection for the V-BLAST systems.

3.2 The Clonal Selection Detection Algorithm for the V-BLAST System

Let us assume a V-BLAST system equipped with M transmit antennas and N receive antennas operating with a coherent binary phase-shift keying (BPSK) modulation, and L is the frame length of transmitted bits.

B is the transmitted bits matrix whose dimension is $M \times L$ during a data burst, we define $B = \{ [\boldsymbol{b}_1^{(1)}, \boldsymbol{b}_1^{(2)}, \cdots, \boldsymbol{b}_1^{(L)}] ; \cdots ; [\boldsymbol{b}_M^{(1)}, \boldsymbol{b}_M^{(2)}, \cdots, \boldsymbol{b}_M^{(L)}] \} \; b_m^{(l)} \in \{-1, 1\}$. B is disposed column by column, so the search space of transmitted sequence is 2^M. The detection

of V-BLAST system can be described as a combinatorial optimization problem(***P***): $\min\{f(\boldsymbol{b}):\boldsymbol{b}\in I\}$, $\boldsymbol{b}=[\boldsymbol{b}_1^{(l)},\boldsymbol{b}_2^{(l)},\cdots,\boldsymbol{b}_M^{(l)}], l=1,2,\cdots L,$ I denotes the set of candidate solutions, f is the antibody-antigen affinity function. $\boldsymbol{b}$ is the variant to be optimized, so the length of antibody is M. It does not need the coding operation for $\boldsymbol{b}$ is binary sequence. Since our intension is to obtain a comparable detection performance with ML detector, we define formula (**2**), according to the ML detection formula [6], as the affinity function:

$$f(\boldsymbol{b})=\|r-H\boldsymbol{b}\|^2. \tag{2}$$

where $\boldsymbol{b}$ is the i-th column of transmitted signal matrix, $i=1,2,\cdots L$, r is the corresponding i-th column of received signal matrix ($N\times L$), H is the matrix channel transfer function.

The clonal selection detection algorithm for V-BLAST can be described as follows:

Step 1. Initialization. Set the halt conditions and algorithm parameters: population size n, clonal scale N_c, mutation probability pm, clonal death proportion $T\%$; Initialize $B(0)=\{b_1(0),b_2(0),\cdots,\ b_n(0)\}\in I^n$; The number of generations k =0.

Step 2. Calculate the affinity of population

$$F(k):\{f(\vec{\boldsymbol{B}}(k))\}=\{f(\boldsymbol{b}_1(k)),f(\boldsymbol{b}_2(k)),\cdots,f(\boldsymbol{b}_n(k))\}. \tag{3}$$

Step 3. Judgement of halt conditions. While satisfy the halt condition, stop; if not, go to *step 4*.

Step 4. Clonal Death Operation $T_{\mathrm{d}}^{\mathrm{C}}$ to $\vec{\boldsymbol{B}}(k)$. Get $\vec{\boldsymbol{B}}'(k)$ (father) by substituting the $T\%$ antibodies with high affinity for the $T\%$ antibodies with low affinity.

Step 5. Clonal Operation T_c^{C} *to* $\vec{\boldsymbol{B}}'(k)$. Get $\vec{\boldsymbol{Y}}(k)$ with population size $n\times(N_c+1)$ by cloning each antibody of $\vec{\boldsymbol{B}}'(k)$, the sum of clone times is N_c.

Step 6. Clonal Mutation Operation T_m^{C} *to* $\vec{\boldsymbol{Y}}(k)$. According to the mutation probability p_m, the cloned antibody populations are mutated as follows: $\vec{\boldsymbol{Z}}(k)=T_m^{\mathrm{C}}(\vec{\boldsymbol{Y}}(k))$, all genes of each antibody of $\vec{\boldsymbol{Y}}(k)$ multiple -1 with probability p_m. It is no need to apply mutation operation to father to avoid the loss of good information of father.

Step 7. Clonal Selection Operation T_s^{C} *to* $\vec{\boldsymbol{Z}}(k)$ *(child).* Calculate the affinity of population, and the antibody with the highest affinity is selected to the next generation among all the antibodies of each group including all children of each father and each father itself, so get the antibodies population $\vec{\boldsymbol{B}}(k+1)$ of the next generation. We also get the optimum antibody of the current generation by checking the highest affinity from $\vec{\boldsymbol{Z}}(k)$. $k=k+1$, go to *Step 2*.

4 Computational Complexity

The ML detector [6] has an exponential complexity ($O(P^M)$, P is the modulation stage, M is the number of transmit antennas) for it need to present a full-search to the search space of the transmitted sequences. When M is large, the complexity is striking.

The OSIC detector offers an obviously decreased complexity, but the complexity is still very high ($O(M^4)$) for many times of pseudo inverse [16].

We define that CA-VBLAST detector has a population size n, clonal scale N_c, mutation probability pm, clonal death proportion $T\%$, number of transmit antennas M, frame length of transmitted signals L and runs for ga generations. During one "data burst", a $M \times L$ transmitted data matrix is presented and operated column by column, so the length of coding is M. For a given L, the computational complexity of the algorithm is decided by two elements: one is ga and the other is the computational complexity per generation N_g. Therefore, the complexity will be $O(N_g \times ga)$. Enact $ga = \alpha M$, where α is a constant. The time complexity per generation for the algorithm can be calculated as follows:

The time complexity is $O(n \times M)$ for initialing antibody population and calculating the antigen-antibody affinities for all antibodies; the time complexity for clonal death operation is $O(\frac{n \times (n-1)}{2})$; $O(N_c \times n \times M)$ for clonal operation and clonal mutation operation, and $O\left(\frac{n \times N_c \times (N_c+1)}{2}\right)$ for clonal selection operation. So the worst total time complexity is

$$O(n \times M) + O(\frac{n \times (n-1)}{2}) + O(N_c \times n \times M) + O\left(\frac{n \times N_c \times (N_c+1)}{2}\right). \quad (4)$$

where, N_c and n are set as constants. According to the operation rules of the symbol O, the worst time complexity of one generation for CA-VBLAST is $O(M)$. Thereupon the complexity of the new algorithm is $O(M^2)$. This implies significant computational savings over the ML detector.

5 Simulation Results and Performance Analysis

To certify the effectiveness of the new detector, the performance of CA-VBLAST is evaluated via computer simulations and compared with the detectors based on standard genetic algorithm [17] (GA-VBLAST), ML algorithm, MMSE-OSIC algorithm and Golden algorithm. BPSK modulation, the frame length 50 and a flat-fading and quasi-stationary rayleigh channel are assumed. Considering the complexity of ML detector, we set the number of transmitted frames 100. 10 times Monte-Carlo simulations are adopted. We consider the V-BLAST systems with 4×4, 4×8 and 8×8

antenna configurations. The maximum number of generations (ga) is considered as stopping criterion in the simulation of CA-VBLAST and GA-VBLAST.

4×4 V-BLAST

We set ga =20 here. We define population size of GA-VBLAST n **=20,** selection probability 0.4, cross probability 0.6, mutation probability $p_m = 1/M$; CA-VBLAST has a population size n =4, clonal scale N_c =4, clonal death probability T% =50%, clonal mutation probability $p_m = 1/M$.

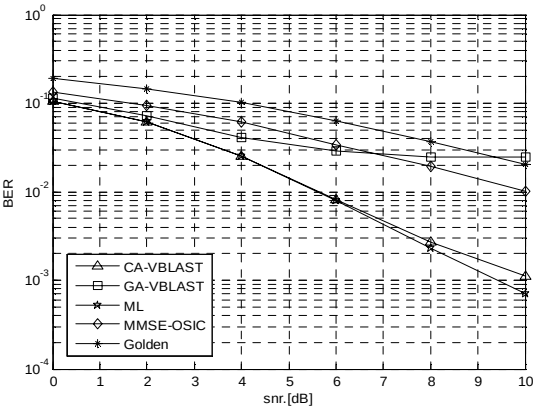


Fig. 2. The BER simulation result versus snr (4×4)

Figure.2 shows that ML detector obtains the optimum BER performance, but its exponential complexity is the highest. CA-VBLAST exhibits a comparable BER with the ML detector and a much better BER than other detectors. BER of CA-VBLAST reaches 10^{-3} when snr =10 dB, and BER equals to 10^{-2} when snr =6dB which saves 4dB than MMSE-OSIC. CA-VBLAST also has a much lower computational complexity than ML, which demonstrates the effectiveness of the new detector. GA-VBLAST does not have a satisfactory BER performance with the same iterative generations, because GA is subject to converge prematurely for the limitations of its algorithm mechanism.

4×8 V-BLAST

CA-VBLAST and GA-VBLAST run for the same generations in 4×8 V-BLAST as in 4×4 V-BLAST and the rest algorithm parameters are all the same.

Figure.3 shows that CA-VBLAST obtains a better BER performance in 4×8 ($M < N$) V-BLAST than in 4×4 ($M = N$) V-BLAST with the same iterative generations. It is obvious that CA-VBLAST has an absolute identical BER compared with ML in 4×8 V-BLAST. The BER of CA-VBLAST equals to 10^{-4} when snr =6 dB

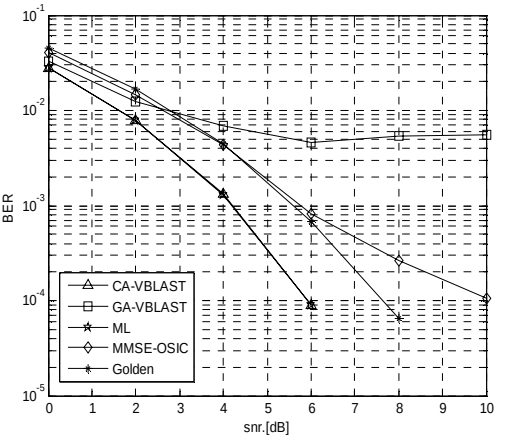


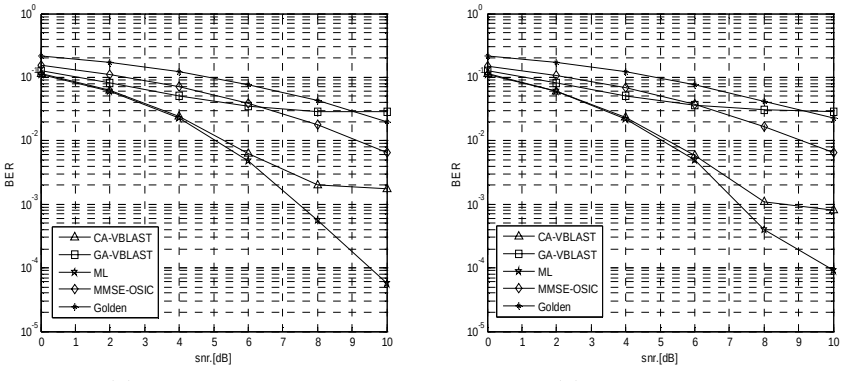
Fig. 3. The BER simulation result versus snr (4×8)

which means a much better improvement compared with 10^{-2} in 4×4 V-BLAST. The conclusion is that the more the number of transmit and receive antennas, the better performance is obtained in V-BLAST system.

8×8 V-BLAST

As M is increased, the solution space is extended, so we should alter the parameters of CA-VBLAST and GA-VBLAST. The V-BLAST system runs for 200 and 400 generations respectively, and the rest algorithm parameters of CA-VBLAST and GA-VBLAST keep invariable.

Figure.4 shows that a just passable BER performance is obtained by CA-VBLAST from 200 runs in 8×8 V-BLAST without altering others algorithm parameters. It is also unsatisfactory from 400 runs. It demonstrates that the number of iterative generations should be increased following with increased number of transmit

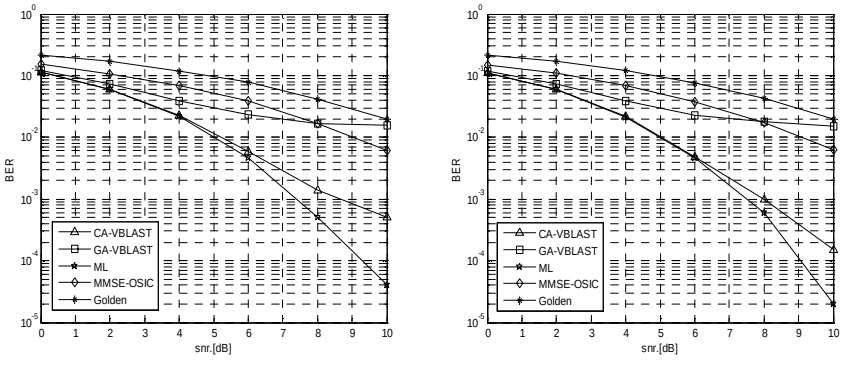


(1) 200 generations (2) 400 generations

Fig. 4. The BER simulation results versus snr (8×8)

antennas, the satisfactory BER performance can not be obtained by increasing the number of iterative generations purely. So the algorithm parameters should be altered to get the comparable BER with ML detector.

For less runs and satisfactory BER, the population size n is increased. The new population size of CA-VBLAST equals to 8 which equals to 4 formerly, and 40 of GA -VBLAST which equals to 20 formerly. The result is followed as Figure.5 (1) from 100 runs. We also increase the mutation probability following with the larger population size for much less runs. The mutation probability p_m is changed from $1/M$ to $2/M$. Figure.5 (2) shows the result from 50 runs.



(1) n =8 $p_m = 1/M$ for CA-VBLAST
n =40 $p_m = 1/M$ for GA-VBLAST
100 generations

(2) n =8 $p_m = 2/M$ for CA-VBLAST
n =40 $p_m = 2/M$ for GA-VBLAST
50 generations

Fig. 5. The BER simulation result versus snr with altered parameters (8×8)

It is obvious that the BER curve of CA-VBLAST in Figure.5 (1) from less runs approaches ML better than in Figure.4. The BER curves of CA-VBLAST and ML are better approached in Figure.5 (2) from 50 runs than in Figure.5 (1). So the conclusion, a much better BER performance should be obtained by increasing the population size and mutation probability from less runs as the transmit antennas are increased, can be presented.

Better population diversity and stronger ability of random search can be obtained from the increased population size and mutation probability, so satisfactory BER performance can also be gained by CA-VBLAST with relatively less iterative generations. But the population size and mutation probability can not be random altered.

6 Concluding Remarks

In this paper, an artificial immune system algorithm for V-BLAST detection inspired by the antibody clonal selection theory was proposed. Theoretical analysis and

computer simulations have shown that the new clonal selection detector could reduce the computational complexity significantly and exhibit better BER than standard genetic algorithm detector and some classical detectors for V-BLAST systems. With proper parameters, the BER curve of CA-VBLAST detector can be obtained which approaches the curve of ML well, so the effectiveness of our algorithm is fully testified.

The future work is that the clonal selection algorithm combined with orthogonal frequency division multiplexing (OFDM) is to be utilized in the V-BLAST system with frequency selective fading channels.

References

1. G J.Foschini and M.J.Gans.: On limits of wireless communications in a fading environment when using multiple antenna. Wireless personal communications, Vol.6, No.3, (1998): 311-335
2. C-N Chuah, G J.Foschini, et al.: Capacity growth of multi-element arrays in door and outdoor wireless channels. Proc of IEEE WCN2000, Chicago, USA. (2003)
3. V.Tarokh, A.Naguib, and N.Seshadri.: Combined array processing and space-time coding. IEEE Trans. Inform. Theory, Vol.45, No.4, (1999): 1121-1128
4. G J.Foschini.: Layered Space-time Architecture for Wireless Communication in a Fading Environment when Using Multi-element Antennas. Bell Labs Technical Joumal, Vol.1, No.2, (1996): 41-59
5. P.W.Wolniansky, G.J.Foschini, G,D,Golden, and R.A.Valenzuela.: V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel. in proc. URSI Int. Symp. Signal, Systems, and Electronics, Pisa, Italy, (1998)
6. A. Adjoudani, E C. Beck, A P. Burg,etc.: Prototype experience for MIMO BLAST over third-generation wireless system. IEEE Journal on Selected Areas in Communications, (2003): 440-451
7. G D. Golden, C J. Foschini, R A. Valenzuela, P W. Wolniansky.: Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture. Electronics Letters, Vol.45, No.1, (1999): 14-16
8. A Bhargave, R. J. P. de Figueiredo, T Eltoft.: A detection algorithm for the V-BLAST system. IEEE Global Telecommunications Conference, (2001): 494-498
9. Castro L N De, et al.: The Clonal Selection Algorithm with Engineering Applications. In: Proc.of GECCO'00, Workshop on Artificial Immune Systems and Their Applications, (2000): 36-37
10. Kim J, et al.: Towards an artificial immune system for network intrusion detection: an investigation of clonal selection with a negative selection operator. In: Proceedings of the 2001 Congress on Evolutionary Computation, (2001): 1244~1252
11. Jungwon Kim, P.J. Bentley.: Towards an artificial immune system for network intrusion detection: an investigation of clonal selection with a negative selection operator. Proceedings of the 2001 Congress on Evolutionary Computation, (2001): 1244~1252
12. Haifeng Du, Licheng Jiao, Sun'an Wang.: Clonal Operator and Antibody Clone Algorithms. Proceedings of the First International Conference on Machine Learning and Cybernetics, Beijing, 4-5 November (2002): 506~510
13. Haifeng Du, Licheng Jiao, Maoguo Gong.: Immune Memorizing Clonal Programming Algorithm. Proceedings of the Thirteenth Chinese Conference on Neural Network. (2003)

14. Haifeng Du, Licheng Jiao, Maoguo Gong.: Immune Clonal Selection Algorithm and Evolutionary Algorithm. Proceedings of the Tenth Annual Science Seminar of Chinese Conference on Artificial Intelligence. (2003)
15. Maoguo Gong, Haifeng Du, Licheng Jiao, Ling Wang.: Immune Clonal Selection Algorithm for Multiuser Detection in DS-CDMA Systems. Australian Conference on Artificial Intelligence, (2004): 1219-1225
16. Wong Kwan Wai, Chi-Ying Tsui, Cheng R.S.: A low complexity architecture of the V-BLAST system. Wireless Communications and Networking Conference, WCNC. IEEE, Vol.1, (2000): 310-314
17. Xiaoping Wang, Liming Cao.: Genetic Algorithm－Theoretical Application and Software Implement. Xi'an Jiao Tong University Press, (2002) (in chinese)

JSCC Based on Adaptive Segmentation and Irregular LDPC for Image Transmission over Wireless Channels*

Rui Guo and Ji-lin Liu

Department of Information Science Electronic Engineering Zhejiang University, 310027, Hangzhou, China
hbsyrgr@yahoo.com.cn

Abstract. In order to improve error resilient capability and transmission efficiency for image transmission over wireless channels, a joint source channel coding (JSCC) scheme was proposed. Adaptive segmentation can segment an image into different size of block with different level of significance; high degree bit nodes within an irregular low-density parity check (LDPC) code can provide unequal error protection (UEP) scheme. So the adaptive segmentation and irregular LDPC coding were combined together to provide JSCC. Simulation results show that the proposed scheme can support robust image transmission in a very effective way with high visual quality.

1 Introduction

JSCC has been proven to be very promising to provide reliable image transmission while maintaining high transmission efficiency[1]. An important means to achieve JSCC is to partition source information with different levels of significance so that different levels of error protection may be applied as just required[2-3]. However, in this paper, we proposed an adaptive segmentation method in terms of pixel intensity variation, which is further incorporated with an UEP scheme to achieve high error resilience for image transmission.

According to the weight of columns in the parity check matrix being equal or not, LDPC code can be divided into two classes: regular code and irregular code. The performance of irregular code is much better than the regular one[4-5]. Except the unique performance to Shannon limit, irregular LDPC code can provide UEP by different degree bit nodes. Accordingly, a robust and efficient image transmission scheme based on adaptive segmentation and irregular LDPC is contrived in this paper.

2 Adaptive Image Segmentation

A non-stationary image source S may be considered as a combination of a set of stationary sub-sources S_1, S_2, S_3 ... S_L, and each sub-source can be considered as stationary. So motivated by this theory, an adaptive segmentation was contrived in

* Work was supported by grant NO.60534070 from the national foundation of China.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 412 – 415, 2006.

this paper, which follows the variation of stationarity. And is determined by the pixel variance σ of the block, which is calculated by the following.

$$\sigma = \frac{1}{M \times N} \sum_{i=1}^{N} \sum_{j=1}^{M} [(x_{i,j} - \bar{x})^2], \qquad \bar{x} = \frac{1}{M \times N} \sum_{i=1}^{N} \sum_{j=1}^{M} x_{i,j} \tag{1}$$

M , N is width and height of each image block respectively.

The algorithm of adaptive segmentation consists of the following steps [6]:

Step 0. Set a threshold for variance control σ_T in terms of certain visual quality guarantee.

Step 1. Select an initial seed block with a size of 8×8 (n is equal to 1)

Step 2. Extend the seed block to surrounding blocks ($n = n + 1$).

Step 3. Calculate the pixel variance σ of the extended block. If the variance within the block is less than the threshold σ_T , set the extended block as a seed block and go to step 2 or go to step 1 otherwise.

3 Proposed JSCC Scheme Based on Segmentation and Irregular LDPC

An ensemble of irregular LDPC code's column and row weight distribution can be described by two polynomials:

$$\lambda(x) = \sum_{i=1}^{d_v} \lambda_i x^{i-1} \,.\, \rho(x) = \sum_{i=1}^{d_r} \rho_i x^{i-1} \tag{2}$$

Where, λ_i and ρ_i represent the percentage of edges emitting from bit nodes and check nodes with degree i ; d_v and d_r are the maximum degrees for bit nodes and check nodes, respectively.

The decoding procedure of irregular LDPC codes is the same as that of regular LDPC codes using an iteratively probabilistic algorithm [7]. In the process of decoding, the bit nodes with high degree can get more information from neighbor check nodes, so the bit nodes with high degree can be decoded very fast with high validity. That is the inherent UEP property of irregular LDPC codes.

The algorithm of the UEP scheme in proposed JSCC consists of the following steps:

(1) Set variance threshold σ_1 and σ_2 ,which satisfy $\sigma_T > \sigma_1 > \sigma_2$.

(2) Compare the pixel variance of the each block σ which has been provided by SSI from adaptive segmentation with σ_1 and σ_2 .

(3) If $\sigma_T > \sigma > \sigma_1$, these blocks (denoted as B_1) need the most powerful error protection and correspond to the bit nodes with highest degrees in irregular LDPC codes.

(4) If $\sigma_1 > \sigma > \sigma_2$, these blocks (denoted as B_2) need a relative lower error protection than blocks B1, corresponded to the bit nodes with moderate degrees in irregular LDPC codes.

(5) If $\sigma < \sigma_2$, these blocks (denoted as B_3) need the lowest error protection and encoded by the bit nodes with lowest degrees in irregular LDPC codes.

Arranging the image block with the above rule, the most important bits are transmitted by the part with the most powerful error protection.

4 Simulation Results

In experiment, we compare two kinds of transmission schemes. The first one is adaptive segmentation with UEP by code1 scheme. The second is JPEG coded image with Equal Error Protection (EEP) by code 2. A 256×256 gray image Lena is used for test. In experiment, SUI-3 channel model was used. BPSK was choosen. The threshold σ_T is equal to 10, σ_1 is equal to 6 and σ_2 is 3 .The interactive number is equal to 50 in the LDPC decoding. Code 1 is irregular LDPC code, the code's block is 8192 bits, and code rate is 1/2, the column polynomial and row polynomial are as fellows:

$$\lambda(x) = 0.1322x + 0.124x^2 + 0.3306x^{19} + 0.4132x^{49} \quad \rho(x) = 0.8678x^{14} + 0.1322x^{15}. \tag{2}$$

Code 2 is regular LDPC code each bit node has 10 edges and each check node has 20 edges. Simulation of different PSNR (peak-signal-noise-ratio) to SNR are presented in Fig.1. As a special example, the PSNR and visual characteristic of reconstructed images transmitted by two schemes when Eb/No=1.6 and 1.1dB are given in Fig.2.

It is clear that the proposed JSCC with adaptive segmentation combined with irregular LDPC outperforms the traditional JPEG coding scheme with equal error protection (EEP) substantially. The image quality of the UEP scheme (PSNR=49.02 dB) is about 17.78 dB higher than that of the EEP scheme (PSNR=31.24 dB) when Eb/No=1.6dB. This is because, channel errors likely corrupt those less significant

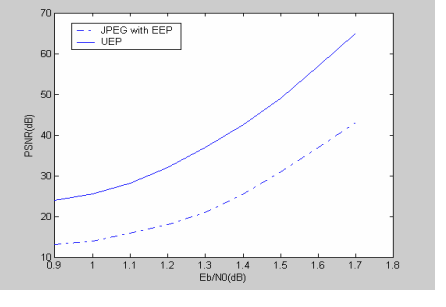


Fig. 1. Comparison of reconstructed

(a)(PSNR=31.24) (b) (PSNR=20.59) (c)(PSNR=49.02) (d)(PSNR=32.94)

Fig. 2. Comparison of Reconstructed image (a) (b) JPEG with EEP, (c) (d) adaptive segmentation with UEP;(a)(c) Eb/N0=1.6dB, (b(d) Eb/N0=1.1dB

blocks, which are relatively easy to be concealed due to less details being included. The most significant blocks, which are coded by the high degree bit nodes in LDPC codes, can get higher error protection.

5 Conclusions

In this paper, an efficient and reliable image transmission scheme based on JSCC is contrived. In this scheme, adaptive segmentation and irregular LDPC coding are designed jointly. Adaptive segmentation potentially leads to a high compression ratio whilst indicating different levels of source significance of individual blocks. Consequently, irregular LDPC code is employed to provide UEP to different important image blocks in terms of SSI. Simulation results confirm that this scheme can support robust image transmission in a very effective way while yielding high visual quality in various channel conditions without any cost.

References

1. H. Zheng and K. J. R. Liu: The subband modulation: A joint power and rate allocation framework for subband image and video transmission. IEEE Trans. Circuits Syst. Video Technol., vol. 9, Aug.(1999) .pp. 823–838
2. Jie Song and K. J. Ray Liu: Robust progressive image transmission over OFDM systems using space-time block code. IEEE Trans.on multimedia, vol. 4, Sep(2002) . pp. 394–406
3. A. Aydın Alatan, Minyi Zhao, and Ali N. Akansu: Unequal Error Protection of SPIHT Encoded Image Bit Streams. IEEE Journal on Selected Areas in Communications, Vol. 18, June (2000) . No. 6
4. D.J.C.Mackay, R.M.Neal: Near Shannon Limit Performance of Low-Density Parity-Check Codes. Electronic letters, Vol.32, (1996) . pp. 1645-1646
5. S. Y. Chung, G. D. Forney, et al: On the design of low density parity check codes within 0.0045 dB of the Shannon limit. IEEE Commun. Letters, (2001). pp. 58-60
6. Yingjun Su, Jianhua Lu,Jing Wang and Jun Gu: Joint source-channel coding for image transmission over noisy channels. Communication Technology Proceedings, 2000. WCC - ICCT 2000. International Conference on Volume 1, 21-25 Aug. vol.1 (2000). pp.136 - 139
7. Ben Lu, Guosen Yue, and Xiaodong Wang: Performance analysis and Design optimization of LDPC-coded MI MO OFDM Systems. IEEE Trans. on Signal Processing. Theory, (2004). pp. 348–361.

Relay-Bounded Single-Actor Selection Algorithms for Wireless Sensor and Actor Networks

ZhenYang Xu[1], Jie Qin[2], GuangSheng Zhang[1], and WenHua Dou[1]

[1] Computer Institute, National University of Defense Technology, Changsha Hunan, 410073, China
[2] School of information Science and Engineering, Henan University of Technology, Zhengzhou Henan, 450052, China
xuzy_sun1@163.com, qinjie0160@yahoo.com.cn, gszhang@nudt.edu.cn, douwh@vip.sina.com

Abstract. Wireless sensor and actor networks (WSANs) are composed of a large number of sensors and a small number of (mobile) resource-rich actors. Sensors gather information about the physical phenomenon, while actors take decisions and then perform appropriate actions upon the environment. Real time and network lifetime are important factors of WSANs. So in this paper, a single-actor selection problem for WSANs is addressed from real time and nodes' Relay Bound constraints first, and then a multi-objective programming is provided. After that, two approximate algorithms, Global Relay-Bounded and MIN-MAX Hops (GRBMMH) and Distributed Relay-Bounded and MIN-MAX Hops (DRBMMH), are put forward. In the performance evaluation, those algorithms are compared with MECT (Minimum Energy Cost Tree) and MPLCT (Minimum Path Length Cost Tree) algorithms.

Keywords: Wireless Sensor and Actor Networks, Real-Time Communications, Energy Efficiency, Relay-Bound.

1 Introduction

In the wireless sensor and actor networks (WSANs), there are a small number of actor nodes as well as a large number of sensor nodes. As in the wireless sensor networks (WSNs), sensor nodes, which are passive elements sensing from the environment, have limited energy, processing and wireless communication capabilities. Those actor nodes, which are active elements acting on the environment, have higher processing and communication capabilities, less constrained energy resources (longer battery life or constant power). Compared with sink nodes in WSNs, actor nodes are closer to sensor nodes around the event area, so the delay of data delivery is shorter, and the real-time requirements of the applications are met easily. There are many potential applications of WSANs, such as battlefield surveillance and microclimate control in buildings, nuclear, biological and chemical attack detection, industrial control, home automation and environmental monitoring [1] [2].

There are single-actor and multi-actor models in WSANs. In the single-actor model, sensor readings may be sent only to one actor. In this way, if the actor has enough

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 416 – 427, 2006.

capability to perform the action, it can immediately perform the action, especially for single-actor task. This implies that in the model latency between sensing and acting may be low as well as there may be no needed for actor–actor coordination [1]. In the multi-actor model, multiple actors can also receive the information about the sensed event. In the model, actor–actor coordination is always necessary; on the one hand each actor receiving sensor data has partial information about the overall event, on the other hand it can avoid many actors making repetitive and redundant decisions to the same event. So the model may cause high network load as well as high latency.

In the paper [3], a distributed coordination framework for WSANs is provided. It considers multi-actor single task model mainly, and involves multi-actor selection algorithm. Some real-time and energy-efficient routing [4] [5] [6] are concerned. Multi-sink position [7], placement of integration points between the wireless and wired network [14], minimum energy data gathering problem with multiple sinks [15], and data acquisition with voronoi clusters [16] are concerned. But single-actor selection problems haven't be presented.

In WSANs, sensor nodes have limited transmission range, and therefore are forced to route their data over multi-hops until the data reach the final destination. So there is an unavoidable problem of unbalanced energy dissipation among different nodes: the nodes, closer to the actor as a collector, are burdened with a heavy relay traffic load (i.e., the "hot spot" problem). The problem leads to the situation where some nodes lose energy at a higher rate and die much faster than others, possibly reducing sensing coverage and leading to network partitioning. So we first introduce a concept: *Relay Bound*. There are two reasons for introducing it: there exists a capacity problem when a node's processing speed becomes the bottleneck; it is necessary for avoiding relaying packets superfluously for one event. Via setting *Relay Bound* threshold, any one node couldn't consume more energy for one event, and the network lifetime of WSANs is prolonged. Besides energy constraint, real-time requirement of the applications is important for WSANs. To satisfy the real-time constraint at the same time, we minimize maximum hops from sources to the selected actor, and put forward multi-objective programming of single-actor selection problem.

In this paper, we investigate schemes to efficiently select single actor in WSANs. Our key contributions are:

- We formulate the single-actor selection problem from network lifetime and real-time aspects. Our algorithms aim to minimize the maximum number of hops from sources to the selected actor while guaranteeing nodes' Relay-Bounded constraint.
- We put forward two approximate algorithms, Global Relay-Bounded and MIN-MAX Hops (GRBMMH) and Distributed Relay-Bounded and MIN-MAX Hops (DRBMMH).We demonstrate the efficiency of the algorithms through simulation.

The paper is organized as follows. In Section 1, we state single-actor selection algorithm from real-time and nodes' Relay Bound. In Section 2, we provide the preliminary model and concepts used in the paper, and propose a multi-objective programming formulation for RBMMH in section 3. We describe a global approximate algorithm GRBMMH and a distributed approximate algorithm DRBMMH respectively in section 4 and 5. Detailed comparative performance evaluation and simulation results are presented in Section 6. Finally, in Section 7 we draw the main conclusions.

2 Network Model and Definitions

2.1 Network Model

The single-actor selection problem, in its simplest form, is to select only one actor as the collector that can receive sensor readings in real-time and efficient way.

The network of sensor and actor nodes is represented as a directed graph G (V, E), where V is a finite set of sensors and actors (vertexes) in a finite-dimension terrain, with $v=|V|$, and the arc set E stands for valid communication links. Let A represent the set of actors, with $m=|A|$. Let N represent the set of sensors, with $n=|N|$. We refer to an actor that is collecting traffic from one or more sources as a collector. Let S be the set of traffic sources, with $s=|S|$. This set represents the sensor nodes that detect the event, i.e. the sensors residing in the event area. It is obvious, $v=n+m$, $V=N\cup A$, $S\subset N$, and $N\cap A=\phi$.

In the paper, we assume: Sensors have the same capability such as communication range, energy, and sensing range. Sensors and actors are immobile.

2.2 Definitions and Decisions

Before providing the multi-objective programming of single-actor selection, we first introduce some symbols and definitions [11]. We don't consider data aggregation.

Definition 1. A *path* from a sensor $i\in S$ to an actor $a\in A$ is a non-empty subgraph $P_{i\to a}$ of G, where $P_{i\to a}=(V_{i\to a}, E_{i\to a})$, $V_{i\to a}=\{i, i_1, i_2, \ldots, i_n, a\}$, $i, i_1, i_2, \ldots, i_n\in N$, $E_{i\to a}=\{(i, i_1), (i_1, i_2), \ldots, (i_{n-1}, i_n), (i_n, a)\}\subseteq E$. The node $i\in S$ is called as the initiator node or traffic source, and the nodes $i, i_1, i_2,\ldots, i_n\in N$ are called as intermediate nodes or relay nodes.

Definition 2. The *cost of an arc* $(i, j)\in E$, e_{ij} is defined to be a real-valued function e: $A\to\Re$. The *cost of a path* $P_{i\to a}$ from a sensor node $i\in S$ to an actor node $a\in A$ is given by

$$e(P_{i\to a})=\sum_{(j,k)\in E_{i\to a}} e_{jk} .$$

Definition 3. The *minimum cost path* $P_{i\to a}^{\min}$ from a sensor node $i\in S$ to an actor node $a\in A$ is a path with the vertex set $V_{i\to a}^{\min}$, where $e(P_{i\to a})$ is a minimum, i.e.,

$$P_{i\to a}^{\min}=\arg\min_{P_{i\to a}}\{e(P_{i\to a})\} .$$

The *minimum cost tree* $T^{\min}=(V', E')$ is a subgraph of $G=(V, E)$, $V'\subseteq V$, $E'\subseteq E$, where

$$T^{\min}=\bigcup_{i\in S}\min_{a\in A}\{P_{i\to a}^{\min}\}.$$

The tree $T_a^{\min}=(V'', E''), a\in A$ is a subgraph of $G= (V, E)$, $V''\subseteq V$, $E''\subseteq E$, where

$$T_a^{\min}=\bigcup_{i\in S} P_{i\to a}^{\min}$$

The tree $T_a^{\min}$ is a collection of all the minimum cost paths from the sensors residing in the event area to the actor *a*. It is worth noting that not all sensors are the vertexes of the minimum cost tree.

Definition 4. Let $s=|S|$, $n=|N|$, $m=|A|$, $v=|V|=n+m$. The path matrix $M_P = \left(p_{jk}^{ia}\right)_{s\times m\times n\times v}$ of the tree $T_a^{\min}$ $a\in A$ is defined by

$$p_{jk}^{ia} = \begin{cases} 1 & \text{if } (j,k)\in E_{i\to a}, i\in S \\ 0 & \text{otherwise} \end{cases}.$$

The elements of the form p_{jk}^{ia} should be read as follows: If we consider a path from a sensor *i* to an actor *a*, and if the link connecting the nodes *j* and *k* lies on that path, then the value of the element is 1, otherwise 0.

Definition 5. Let $i\in S$, $a\in A$. The *set of relay nodes* of the path $P_{i\to a}$ is defined as

$$R_{i\to a} = \{j\in N : j\in V_{i\to a}-\{a\}\}\subseteq \mathrm{N}.$$

Definition 6. The *hops* of the path $P_{i\to a}$ are defined as

$$l_{i\to a}=|R_{i\to a}|$$

Theorem 1. Let $P_{i\to a}$ be the path from a sensor $i\in S$ to an actor $a\in A$. Then we have

$$l_{i\to a} = \sum_{j\in N}\sum_{k\in V} p_{jk}^{ia}.$$

The result of theorem 1 can be used to calculate the hops of each path in the tree T_a.

Definition 7. Let $a\in A$. The *branch set of a relay node* $j\in N$ is defined as

$$B_j^a = \{i\in S : j\in R_{i\to a}\}.$$

Similarly, the *set of branch nodes or* the *branch set of an actor node* $a\in A$ is defined as

$$B^a = \bigcup_{j\in N} B_j^a.$$

Definition 8. The *branch size of a relay node* $j\in N$ is defined as

$$b_j = \sum_{a\in A}\left|B_j^a\right|.$$

Theorem 2. Let $T^{\min}$ be the minimum cost tree in a sensor and actor network, and $j\in N$ be an arbitrary sensor node. Then, we have

$$b_j = \sum_{a\in A}\sum_{i\in N}\sum_{k\in V} p_{jk}^{ia}.$$

Definition 9. The *relay bound* threshold *C* is defined as the maximum packets allowed to be relayed by any sensor between the instant when the sensors sample the physical features of the event and the instant when the actor receives a data packet describing these event features.

In this study, we define the network lifetime as the time duration till the first sensor node runs out of initial energy. The threshold *C* can be calculated before network is deployed.

Given initial energy *E*, network lifetime *T*, event frequency *f*, maximum energy dissipation of receiving and relaying a packet *E1* (it is related to communication radius and the deployed environment), then the relay bound C of the sensors is

$$C \le \lfloor E / (E1*T*f) \rfloor .$$

3 Multi-objective Programming

Network lifetime and real-time are important factors for single-actor selection problem. So we model the single-actor selection problem as multi-objective programming from relay bound and the minimum of maximum hops. We introduced the notation $\boldsymbol{h_i}$ that is used in the problem formulation.

-$\boldsymbol{h_i}$ is a binary variable representing actor *i*, that equals 1 iff actor *i* is selected as a collector.

Given: $\boldsymbol{C}$. (1)

Find: h_k , p_{jk}^{ia} . (2)

$$\min \max_{a \in A} \sum_{j \in N} \sum_{k \in V} p_{jk}^{ia} \ . \tag{3}$$

Subject to:

$$\sum_{a \in A} \sum_{i \in S} \sum_{k \in V} p_{jk}^{ia} \le C \qquad \forall j \in N \tag{4}$$

.

$$\sum_{k \in V} p_{ik}^{ia} - \sum_{k \in V} p_{ki}^{ia} = h_a \quad \forall a \in A, \forall i \in S \ . \tag{5}$$

$$\sum_{k \in V} p_{jk}^{ia} - \sum_{k \in V} p_{kj}^{ia} = 0 \ \ \forall i \in S, \forall a \in A, \forall j \in N\text{-}S. \tag{6}$$

$$\sum_{k \in V-A} p_{ka}^{ia} - \sum_{k \in V-A} p_{ak}^{ia} = h_a \quad \forall i \in S, \forall a \in A \ . \tag{7}$$

$$\sum_{a \in A} h_a = 1 \qquad\qquad h_a \in \{0,1\} . \tag{8}$$

The objective function (3) minimizes the maximum hops of the network. The constraint (4) expresses that the relay nodes of each sensor node from sensors in the event area to the selected actor are less than the relay bound threshold *C*. Constraints (5-7) express conservation of flows, i.e. each source generates a flow, which is collected by an actor. In particular, constraint (5) guarantees that a source node

generates a flow on the tree of the selected actor, and only on that one; while non-source nodes do not generate any flow. Constraint (6) imposes that the balance between incoming and outgoing flows is null for non-source and non-actor nodes. Constraints (7, 8) require that only one actor collect flows generated by each source.

Integer multi-objective programming belongs to a class of problems known as NP-hard. It is difficult to solve the problem even with the professional solver such as cplex[8]. So a global approximate algorithm of relay-bounded and min-max hops is provided in next section.

Input: set of sources S, set of sensors N with relay bound threshold C, set of actors A, graph G on the set N∪A with capacities on its vertices.
Output: an actor selected as the collector.
BEGIN
for all $a \in A$ do
Set MaxHop(a)=inf; Set TotalEnergy(a)=inf;
end for
while (A not empty) do
Select an actor a from A.
Let G' be the subgraph of G induced on $N \cup \{a\}$, with the same capacities as G.
Flag=true;
for each sensor $j \in S$ do
Try to find a path $P_{j \to a}$ with minimum hops from j to a. (the path with minimum energy dissipation is selected if there exist multi-path)
if No path is found then
Flag=false; break;
end if
for all $k \in P_{j \to a} \cap N$ do
$C_k = C_k - 1$;
if $C_k = = 0$ then
Remove the node k and the edges related to k from G'.
end if
end for
end for
if (Flag) then
Set MaxHop(a)←maximum number of hops;
Set TotalEnergy(a) ←total energy consumption per event.
end if
$A = A - \{a\}$;
end while
The actor with Minimum value in MaxHop is selected as the collector. (If exist multi actor with same MaxHop, the actor with minimum TotalEnergy is selected.)
END

Fig. 1. A skeleton of GRBMMH

4 A Global Approximation Algorithm of RBMMH

In the section, we provide a Global Relay-Bounded and Minimum Maximum Hops algorithm (GRBMMH) for single-actor selection. In GRBMMH, sensors try to find an actor so that at the premise of relay bound constraint, the minimum hops from the sensors in the event area to that actor can be constructed and thus event information can be transmitted in time.

We design the algorithm as follows. We iteratively consider the maximum hops from sources to each actor. For computing the maximum hops, we iteratively pick a source node, and find a path with minimum hops to the actor. If there are multi-path, the path with minimum energy dissipation is selected. Since there is the relay bound in every node, the graph is reconstructed after one path is found. For each node in the path, the relay bound is decreased one. If the relay bound is zero, the node and the edges related to it are removed from the graph. At last, the maximum hops of every actor are calculated and the actor whose maximum hops is smallest is selected as the collector. Figure 1 shows a skeleton of the algorithm.

5 A Distributed Approximation Algorithm of DRBMMH

Before providing the algorithm, we introduce some assumption. Each sensor node is aware of: i) its position, as the sensor node can be equipped with a GPS receiver, or the position can be determined by means of localization techniques [8] [9]; ii) the position of its neighbors, as every node periodically sends its position to its neighbors; iii) the position of the actors, as each actor periodically beacons its position in the sensor field; iv) the network is synchronized by means of one of the existing synchronization protocols [16].

The shorter the distance between the center of event area and the selected actor is, the faster the actor may take appropriate action. Moreover, if all the actors have same action range, it is more possible that the event area is fully covered by the closest actor's action range; Even if different actor has different action range, the actor is closer to the actors whose action range cover the event area, which is help to decrease time-delay of action.

So in our distributed algorithm, the main idea is that the actor which is closest to the center of the event is selected as the collector, and that the network, from source nodes to the actor, is constructed to transfer sensor reading. In the tree framework rooted at the actor, the node, which is closest to the collector, has more heavy traffic. The number of its branch nodes will be over the relay bound threshold. So the tree structure is unable to meet our requirement. At first each source in the event area sends a REQUENT packet to the node with best forward in random time. When receiving a REQUENT packet, the node decides whether the number of its branch nodes is over the relay-bounded threshold. If not, the node sends an ACK packet for acknowledge, and else sends a UNACK packet to the sender. For the node receiving the UNACK packet, it selects the node with second closer to the actor as the relay node, and sends REQUEST packet again. Since in the distributed algorithm, we only consider the case that for each source node, there is a greedy path to the actor, and each node in the path meet the relay threshold, so the traffic network could be

constructed via REQUENT/ACK/UNACK packets. For every event source, each node records the id of its relay node and the number of packets to each relay node. When an event happens, source nodes transfer the reading to their relay node. According to the number of packets to their relay node, intermediate node sends the receiving packets to corresponding relay node.

6 Performance Evaluation

Before evaluating the performance of those algorithms, we introduce three metrics:

Average number of hops: without collision and retransmission, average hops reflect the time by the packets traveled. The smaller average number of hops is, the shorter the delay of data delivery is.

Average energy per packet: This metric represents the average energy consumed in the network for transmitting and relaying a data packet until the actor successfully receives it.

***Maximum packets relayed by the nodes*:** this metric represents network lifetime from one aspect. The metric is bigger; it means that there exists one node's energy consumed quickly, so the network lifetime is shorter.

To contrast the algorithms of GRBMMH and DRBMMH, we introduced two algorithms: minimum energy cost tree (MECT) and minimum path length cost tree (MPLCT). In MECT, sensors try to find an actor so that the minimum energy cost path from the sensors in the event area to that actor can be constructed and thus event information can be transmitted in an energy efficient way. In MPLCT, sensors try to find an actor so that the minimum path length cost tree from the sensors in the event area to that actor can be constructed. The two algorithms are implemented based global network status also.

All algorithms were implemented in Matlab [13]. We consider three different simulation scenarios. In each scenario, there is a square area of 100mX100m, and circle event area with radius 20m and centre coordinate (50, 50). Four actors are randomly placed at 10mX10m area in the four different corner of the deploy area. In scenario 1: we vary sensors nodes with the number ranging in [50,300], step 25. In scenario 2: 200 sensors nodes are randomly deployed. We vary Communication radius of the sensors from 20m to 36m. In scenario 3: 200 sensors nodes are randomly deployed. We vary relay threshold from 4 to 10 for the approximate algorithms. For different setup in each scenario, we simulate at least 50 times. In the simulation, we adopt the energy model: $e_{ij} = \kappa d_{ij}^{\alpha} + \tau$ [12] (where d_{ij} is the Euclidean distance between nodes i and j.), $\alpha = 4$, $\tau = 50$nJ/bit, $\kappa = 100$pJ/bit/m$^{\alpha}$. The transmission range of sensors is set to 25m in the scenario 1 and 3. The relay bound threshold is set to 6 in the scenario 1 and 2.

In scenario 1, the study focuses on sensor density effect on different algorithms. Figure 2 (a) we can see that the average hops of MECT are most, are 9 when sensor number is more than 100. The average hops of other algorithms are almost same, and less than 4. Figure 2 (b) shows, the average energy per packet of GRBMMH and

DRBMMH are bigger than that of MPLCT and MECT. Since there are relay threshold in every node, some packets are transferred by far path. For MPLCT, GRBMMH and DRBMMH, average energy per packet change little as sensor density increases. The value of MECT drops slowly. This is since a sensor is likely to select the neighbor node with the nearest distance to relay a packet in MECT. The distance between nodes decreases as sensor density increases. For other algorithms, the minimum hops are considered, so sensor density doesn't influence them. From figure 2(c), the maximum packets relayed increase quickly in MECT as sensor density increases. When the sensor number is 300, the maximum packets relayed by one node are 34. Compared to MECT, the value increases slowly in MPLCT, is 16 when the sensor number is 300. For there are relay threshold in every node, the maximum packets relayed are 6. It means that the network lifetime of GRBMMH and DRBMMH are bigger than that of other algorithms.

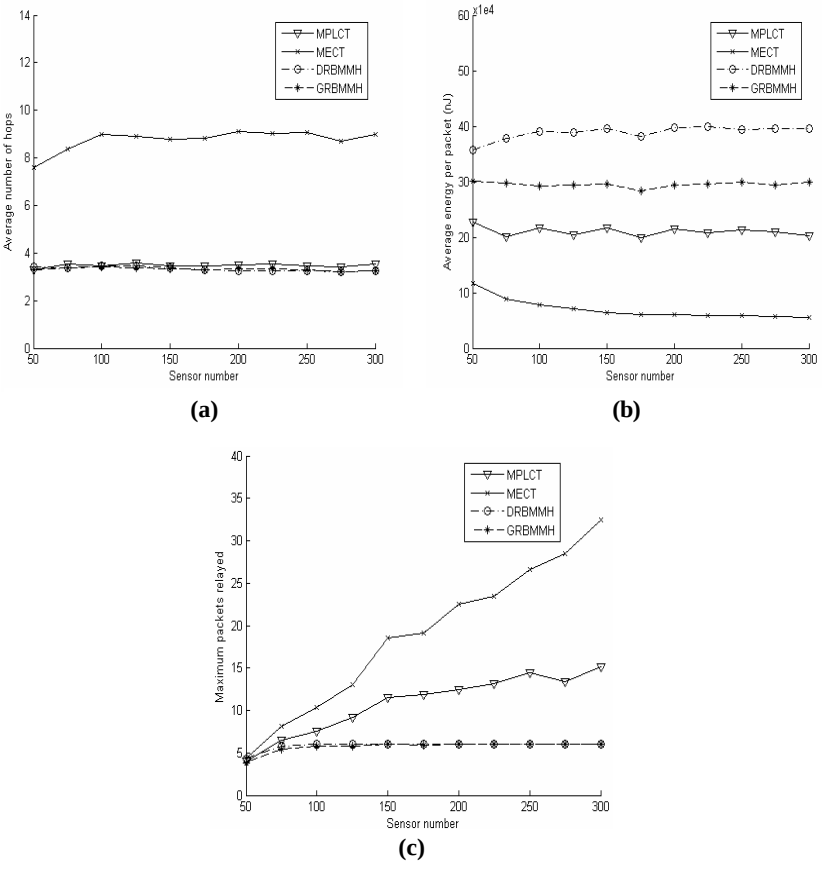


Fig. 2. (a) Sensor density vs. Average number of hops, **(b)** Sensor density vs. Average energy per packet, **(c)** Sensor density vs. maximum packets relayed

In scenario 2, the study focuses on communication radius effect on different algorithms. From figure 3 (a) we can see that the average hops of MECT are most, are 9. The average hops of other algorithms are almost same, and decrease slowly as communication radius increases. Figure 3 (b) shows, for MPLCT, GRBMMH and DRBMMH, average energy per packet change rapidly as communication radius increases. The value of MECT keeps steadily since radius couldn't influence the way of selecting relay node. For other algorithms, a sensor may relay a packet more possible to the neighbor node with the most forward within radius. As communication radius increases, the energy dissipation of single hop increases also, so average energy per packet augments. From figure 3(c), we can see that the maximum packets relayed change a little in MECT as communication radius increases. Compared to

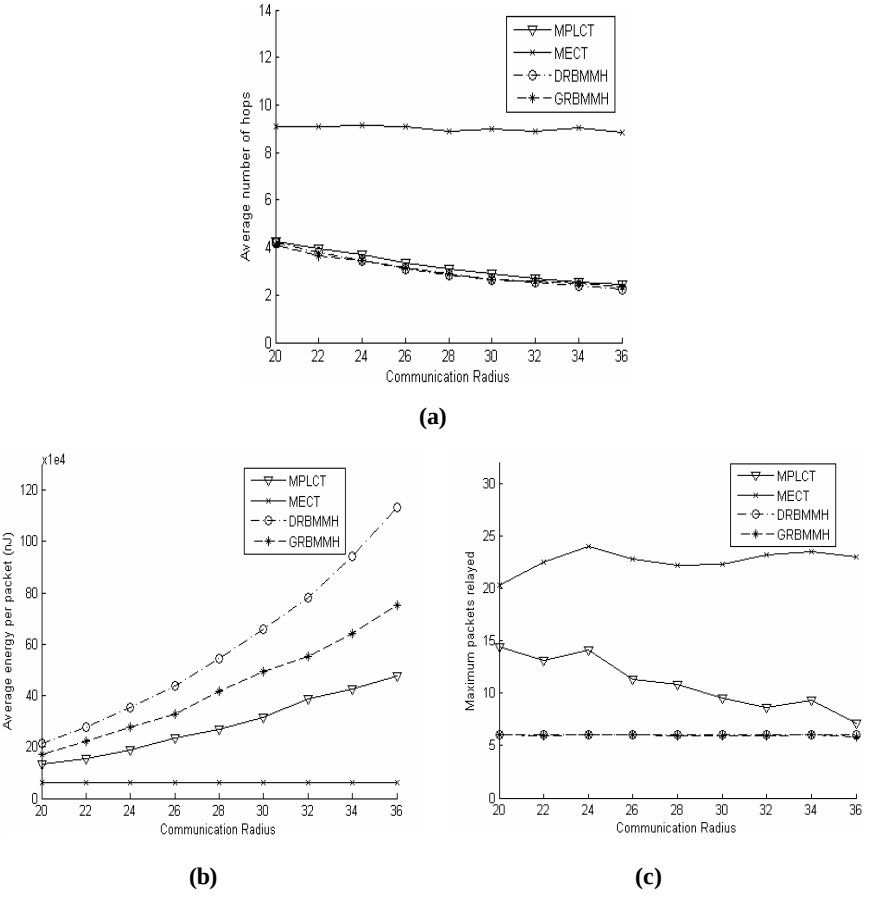
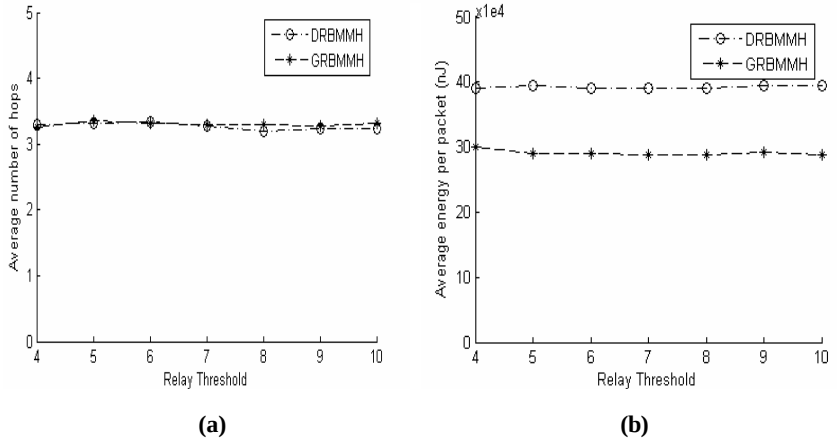


Fig. 3. (a) Communication radius vs. Aver-age number of hops, **(b)** Communication radius vs. Aver-age energy per packet, **(c)** Communication radius vs. maxi-mum packets relayed



(a) **(b)**

Fig. 4. (a) Relay threshold vs. Average number of hops, **(b)** Relay threshold vs. Average energy per packet

MECT, the value drops slowly in MPLCT, is 7 when the sensor number is 300. For GRBMMH and DRBMMH, there are relay threshold in every node, the maximum packets relayed are 6.

In scenario 3, the study focuses on relay threshold effect on approximate algorithms. Figure 4 (a-b) shows, for GRBMMH and DRBMMH, as relay threshold increases, there are no change almost for average energy per packet and average number of hops. So at the premise of meeting real-time, the relay threshold should be decreased as soon as possible, which is help to improve the lifetime of the network.

From the above performance evaluation, although in GRBMMH and DRBMMH average energy per packet are larger than that of MECT and MPLCT, maximum packets relayed by any node are reduced via setting relay threshold. Total energy dissipation is allocated among more nodes, so the network lifetime is prolonged. At the same time, real-time requirement is met via min-max hops.

7 Conclusion

In the paper, we provide a multi-objective programming of single-actor selection problem from network lifetime and real-time aspects, and put forward two approximate algorithms GRBMMH and DRBMMH. The algorithms can prolong the lifetime of the network, and meet real-time requirements of the applications via setting the relay threshold.

In the future, we plan to perfect the algorithms RBMMH from contention based MAC. And then single-actor coordination framework is one of our future works.

References

[1] F. Akyildiz and I. H. Kasimoglu, "Wireless sensor and actor networks: Research challenges," Ad Hoc Networks (Elsevier), vol. 2, no. 4, pp. 351–367, October 2004.

[2] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, "Wireless sensor networks: a survey", Computer Networks (Elsevier) Journal 38 (4) (2002) 393–422.

[3] T. Melodia, D. Pompili, V. C. Gungor and Ian F. Akyildiz, "A Distributed Coordination Framework for Wireless Sensor and Actor Networks", MobiHoc'05, May 25–27, 2005, Urbana-Champaign, Illinois, USA.

[4] T. He, J. Stankovic, C. Lu, and T. Abdelzaher, "SPEED: A real-time routing protocol for sensor networks," in Proceedings of IEEE ICDCS, Providence,RI, USA, May 2003, pp. 46–55.

[5] E. Felemban, C.G. Lee, E. Ekici, R. Boder, and S. Vural, "Probabilistic QoS Guarantee in Reliability and Timeliness Domains in Wireless Sensor Networks," in Proceedings of IEEE INFOCOM 2005, Miami, FL, USA, Mar. 2005.

[6] K. Akkaya, M. Younis, "An Energy-Aware QoS Routing Protocol for Wireless Sensor Networks", Proc. of the IEEE Workshop on Mobile and Wireless Networks (MWN 2003), Providence, RI, May 2003.

[7] H. Kim, Y. Seok, N. Choi,Y. Choi, and T. Kwon, "Optimal Multi-sink Positioning and Energy-efficient Routing in Wireless Sensor Networks" in ICOIN 2005, Jeju Island, Korea, January 31- February 2, 2005, Proceedings.

[8] CPLEX solver. Available: http://www.cplex.com

[9] T. He, C. Huang, B. Blum, J. Stankovic, and T. Abdelzaher, "Range-Free Localization Schemes for Large Scale Sensor Networks," in Proceedings of Mobicom, 2003.

[10] Eylem İlker Oyman, "MULTIPLE SINK LOCATION PROBLEM AND ENERGY EFFICIENCY IN LARGE SCALE WIRELESS SENSOR NETWORKS" doctor paper, Boğaziçi University 2004.

[11] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," IEEE Transactions on Wireless Communications, vol. 1, no. 4, pp. 660–670, Oct. 2002.

[12] Matlab. http://www.mathworks.com

[13] R. Chandra, Lili Qiu, K. Jain, M. Mahdian. "Optimizing the Placement of Internet TAPs in Wireless Neighborhood Networks." ICNP, Oct. 2004.

[14] Kevin Yuen, Baochun Li, Ben Liang, "Distributed Data Gathering in Multi-Sink Sensor Networks with Correlated Sources." to appear in the Proceedings of IFIP Networking 2006, Coimbra, Portugal, May 15-19, 2006.

[15] H. Dubois-Ferriere and D. Estrin, "Efficient and Practical Query Scoping in Sensor Networks," Tech. Rep. 2004-39, CENS/UCLA Tech Report, 2004.

[16] B. Sundararaman, U. Buy, and A. Kshemkalyani, "Clock synchronization for wireless sensor networks: a survey," Ad Hoc Networks (Elsevier), vol. 3, no. 3, pp. 281–323, May 2005.

Probability Based Weighted Fair Queueing Algorithm with Adaptive Buffer Management for High-Speed Network

De-Bin Yin and Jian-Ying Xie

Department of Automation, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, China
{yin_db, jyxie}@sjtu.edu.cn

Abstract. Future high-speed networks will simultaneously support multiple types of services over a single physical infrastructure. Packet scheduling discipline is a crucial technique to enable this. This paper proposes a new weighted fair queuing discipline - Probability based Weighted Fair Queueing, P-WFQ - for high-speed, integrated-service, packet-switched networks. P-WFQ reduces the complexity of implementation by avoiding the main problem of traditional weighted fair queueing algorithm - the calculation of the weight parameter for each packet. It uses a random number to find the next packet to be serviced. In addition, it uses a novel grouping technology, which congregates large number of different flows into a smaller number of groups and reduces the complexity of P-WFQ greatly. This makes it suit for high-speed networks. The simulations prove the validity and practicability of P-WFQ.

1 Introduction

The integration of a wide range of services with different Quality of Service (QoS) requirements in high-speed networks has made the efficient use of resources an issue of great practical and research importance. In packet-switched networks, packets from different sessions belong to different services and administrative classes, which interact with each other when they are multiplexed at the same output link of a switch. The scheduling algorithms at switching nodes play a critical role in controlling the interaction among different traffic streams and different service classes [10].

In the literature, many packet scheduling algorithms have been proposed. Among them, a class of service disciplines called packet fair queueing (PFQ) algorithms have received much attention. PFQ algorithms approximate the idealized generalized processor sharing (GPS) policy [3]. While there are many proposed PFQ algorithms with different tradeoffs between complexity and accuracy [1], [3], [4], [6], [7], [8], few real implementations can be used in high-speed networks while maintaining important properties of GPS (delay bound, fairness and worst-case fairness). The key difficulty is that PFQ algorithms require buffering on a per-session basis and nontrivial service arbitration among all sessions. There are three major costs associated with the arbitration: 1) the computation of the system virtual time function; 2) the management

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 428 – 437, 2006.

of a priority queue to order the transmission of packets; and 3) the management of another priority queue to regulate packets [10].

Compared to GPS-based queueing algorithm, Round-Robin based algorithms have the main advantage - low implementation complexity. So it is more suitable for high-speed networks in terms of the speed. However, conditional RR has a major problem - unfairness, which is caused by possibly different packet sizes used by different flows. On the other hand, RR algorithm can not grantee specific delay bound. In order to improve its performance, many new RR based algorithm have been proposed, such as Weighted Round Robin (WRR) [2], Deficit Round Robin (DRR) [5] and Urgency-based Round Robin (URR) [9]. But in each round-robin, all of these need to calculate weight parameters for all classes once or servral times. Then they are not very suitable for high-speed networks.

In recent years some architectures have been proposed to schedule packets in high-speed networks. In [12] Stoica proposed a core-stateless fair queueing architecture, which uses an island concept to devide routers into edge routers and core routers. Edge routers maintain per-flow state; they estimate the incoming rate of each flow and insert a label into each packet based on the estimate. Core routers maintain no per-flow state; they use FCFS packet scheduling discipline. Stephens used a uniformly exponential spanning technology to devide flows into groups in [10] to meet the requirement of high-speed networks. Inspired by [11], we have proposed a new weighted fair queueing algorithm - P-WFQ, which uses the flows' relative weight parameters as the serving probabilities in one round-robin. In this paper, we extend P-WFQ to high-speed networks. In order to reduce the complexity to meet the speed requirement of high-speed networks, we use a new grouping technology inspired by Stephens [10], which congregate flows into groups according to their weight parameter.

The paper is organized as follow. The next section presents the model of P-WFQ discipline and its implementing algorithm. Section III presents some theoretical results. In section IV, we demonstrate its ability in providing service differentiation as well as fairness among traffic classes through simulations. Finally, in section V the paper ends with the main conclusions.

2 P-WFQ Discipline

As shown in Fig. 1, our P-WFQ scheduling algorithm is composed of Flow Grouping, Adaptive Buffer Manager (ABM) and P-WFQ queue scheduler. The Flow Grouping divides N flows into M groups according to their weight parameters. The ABM automatically adapts the sizes of different flows' buffer according to the number of dropped packets in constant time interval. It optimizes the utilization of server's queue buffer and limits the queueing delay with some packets dropped. In the high-speed networks, ABM is disabled for the consideration of simplicity. The P-WFQ queue scheduler selects the next queue packet to be served on the basis of serving probability. In this paper, we suppose that all packets have the same length. The queue length and buffer length are counted in number of packets.

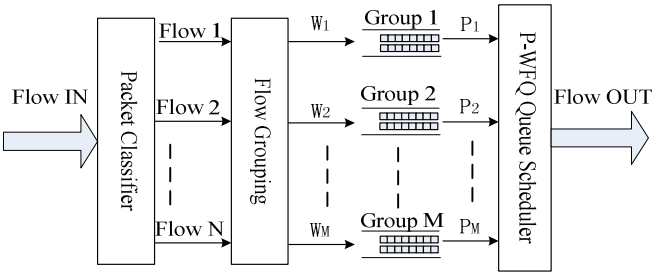


Fig. 1. P-WFQ Scheduling Discipline

The P-WFQ discipline is non-preemptive. Each flow of packets has its own finite queue. Packets in the same queue are served in the First-Come-First-Served (FCFS) fashion. Consider a single-server system. Suppose there be N flows with packets transferred by this switch. Each flow has a weight w_i $(1 \le i \le N)$. So the serving probability function is defined as

$$p_i = \begin{cases} \dfrac{w_i}{\sum_{j \in AQ(N)} w_j}, & i \in AQ(N) \\ 0, & i \notin AQ(N) \end{cases} \tag{1}$$

Where $AQ(N)$ is the set of current active flows. In this discipline we do not use priority conception. All flows are served on the base of its weight parameter.

2.1 Flows Grouping

In order to reduce the overall complexity of PFQ-based algorithms, Bennett proposed a new grouping architecture in [16]. His grouping architecture is based on the calculated virtual start time and virtual finish time. Inspired by this, we propose another simpler grouping algorithm to reduce the number of sessions considered by the scheduler.

Consider a single-server system with N packet flows. Each flow has its own weight parameter w_i. Without loss of generality, supposes there are M different weight parameters. So we can group N different flows into M groups according to flows' weight parameters. Flows with same weight are assembled into the same group. Denote by w_i the weight of flow i, S_j the number of flows in group j, W_j the sum weight of flows in group j and P_j the service probability of group j in one round-robin. Then we have the following:

$$\sum_{j=1}^{M} S_j = N \tag{2}$$

$$W_j = \sum_{k=1}^{S_j} w_k \tag{3}$$

$$P_j = W_j \Big/ \sum_{k=1}^{M} W_k = \sum_{l=1}^{S_j} w_l \Big/ \sum_{x=1}^{N} w_x \tag{4}$$

The P-WFQ scheduler use P_j to select the next group to be serviced. Within groups, flows are serviced in Round-Robin manner. Then the complexity of P-WFQ is reduced from dealing with N flows to dealing with M groups. The next question is that how many different weights maybe exist in the networks. There are only four different QoS classes in IETF's DiffServ model. In real networks, the number of different QoS classes should be limited. So compared to the number of flows N, M is very small. In this paper we only consider high-speed networks with fixed-length packets, like ATM networks.

2.2 P-WFQ Queue Scheduler

P-WFQ queue scheduler uses formula (4) to calculate the serving probability of each group in the current round-robin. It works as follows:

a) The server monitors all flows in the system to inquire about the set of active flows.
b) Serving cycle always starts from $i = 1$, i.e. starting from the first active group. As we have explained before, the first group is the group with the biggest weight parameter (the sum of all active flows' weight parameters in this group) in this server.
c) In every serving cycle, the probability of serving the first packet of queues within the same group is calculated according to formula (4).
d) Only when all queues are empty (no active flow), the server goes to idle status. That is, the server is work-conserving.

The following algorithm outlines a simple implementation of the P-WFQ discipline:

1) After initialization, the server inquires to determine the set of active flows $AQ(N)$. Then calculate the servicing probability parameter p_i for each active queue. Sort queues with p_i. Queues with bigger p are head of those with smaller. Congregate flows into several groups according to their relative weights as we have described above. If all queues are empty (no active flow), the server is waiting.
2) Generate a random number $RN(1)$, $0 \le RN(1) \le 1$. Define a variable *sum* with initial value 0.
3) Set j = the first group in $AQ(N)$, $sum \leftarrow P_j$.
4) If $RN(1) \le sum$, go to step 6. Otherwise go to step 5.
5) Set j = the next group in $AQ(N)$, $sum += P_j$. Go to step 4.
6) Serve the first packet of all queues in group j. Then go to step 1.

3 Analytical Results

In P-WFQ discipline we use a random number $RN(1)$ to find the next packet to be served. So the performance of P-WFQ discipline is largely depended on the random

number generating algorithm. In this section, we begin with the following assumption: the random number generating algorithm is perfect, so the random number is uniformly distributed within (0,1) strictly. Then service probabilities are totally determined by weight parameters.

3.1 Complexity of Implementation

Suppose there are M active flows, which are sorted according to service probabilities from large to small:

$$p_0 \geq p_1 \geq ... \geq p_{(M-1)} \tag{5}$$

Then the average number of floating-number (variable *sum* and random number $RN(1)$) compare operations to find the next packet to be served is

$$Num_{aver} = p_0 + 2 \times p_1 + ... + (M-1) \times p_{(M-2)} + M \times p_{(M-1)} \tag{6}$$

Formula (9) shows that sorting queues with service probabilities from big to small can reduce the number of compare operations within one round robin. This will improve the executing speed of P-WFQ discipline. Suppose there are M active flows in the current system. In the worst case (the first packet of the last active queue – the queue with the smallest weight in the active queues set – will be transferred in this round robin), the discipline needs M floating-point compare operations to find the packet to be sent. So the implementation complexity of P-WFQ is $O(Num_{aver})$. Compared to GPS/Virtual-Time based PFQ disciplines, P-WFQ only has one weight parameter for each active flow, no parameters for each packet. There is no need to calculate relative weight parameters until events of new flow entering the server or current active flow becoming idle happen. So the operation complexity of P-WFQ is greatly less than GPS/Virtual-Time based disciplines. Even compared to RR based disciplines, which are proud of their low implementation complexity, P-WFQ still has comparable performance. This is profited from that P-WFQ does not have to trace with the changes of normalized relative weights for each active flows dynamically.

3.2 Fairness and Delay Bound

All active flows share the server on the basis of their relative weight calculated by formula (1). On the assumption that the random number $RN(1)$ is perfect, we can say that in an enough long time interval, the throughputs gained by two active flows will proportional to their weight. P-WFQ discipline is long-time fair.

Due to the same assumption, each flow will get its weighted bandwidth share regardless of other flows' traffic patterns. So we get the following delay bound. Due to space limitation, their proofs are omitted.

Theorem 3.1. Consider a flow *i* in a system with P-WFQ. Assume traffic arrival of each active flows is (σ, λ) upper constrained, i.e. the number of bits arrived in interval [s, t] is at most $(\sigma + \lambda(t-s))$. Denote by M the number of current active flows in the

system, w_i the weight parameter of flow i, and $L^{\max}$ the maximum packet length. Then, the waiting time in queue of flow i packet is bounded by:

$$0 \le d_i \le \frac{L^{\max}}{C} + \frac{\sigma_i}{C} \times \frac{\sum_{j=1}^{M} w_j}{w_i} - D_i^k \tag{7}$$

Where $\sum_{j=1}^{M} w_j$ is the sum of weights of current active flows, C is the throughput of this server.

4 Simulation Results

We have done many simulations to demonstrate the performance of P-WFQ discipline by using network simulation software ns-2 [18]. For space limitation, only some of them are present here. They are divided into two groups. The first group of simulations is to demonstrate the fairness performance. The second group is to demonstrate the delay bound.

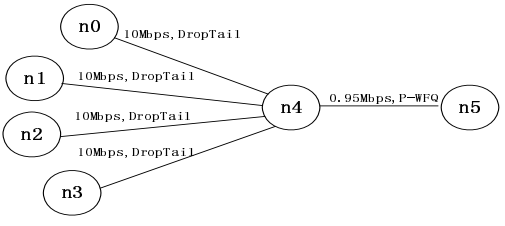


Fig. 2. Network Topology - Fairness Examination

4.1 Service Fairness

This group of simulations is to demonstrate the service fairness of P-WFQ discipline. The network topology used in the simulation is shown in Fig. 2. Node 0, 1, 2 and 3 are linked to node 4 with 10Mbps bandwidth, Drop-Tail discipline respectively. Node 4 is linked to node 5 with 0.95Mbps, P-WFQ discipline. Node 0, 1, 2 and 3 are the data sources. They use the same traffic generator: CBR with 4ms sending interval. The packet buffer of node 4 is 500 packets. It is shared by the 4 data flows equally. Then the buffer length of each flow is 125 packets. The weights of node 0, 1, 2 and 3 are 4, 2, 1 and 1 respectively. In this simulation ABM is not enabled. The simulation time is 100s. Figure 3 shows the simulation result. The bandwidth used by each flow is averaged on every 0.5s. The three vertical lines present the theoretically assigned bandwidth.

From Fig.3 we can see that throughputs of each flows are proportional to their relative weights when the server is overload. This demonstrates the short-term fairness performance of P-WFQ discipline. Because of the probability nature, which is caused by the use of random number, the performance of P-WFQ in small time-scale is not as

good as that in the large time-scale. But in the real networks, a data flow normally last several minutes. In this time-scale we can see that P-WFQ is fair. This is also demonstrated by simulation results shown in table 1.

Using the same network topology and parameters, but replacing the discipline on the link between node 4 and node 5 with DropTail, WFQ and WF^{2}Q+, we get table 1. In this table, "Sent" is the number of packets received by node 4, "Received" is the number of packets received by node 5 and "Percent" is the transferring ratio of this link. Table 1 shows that P-WFQ is same as WFQ and WF^{2}Q+ on the fairness of weighted bandwidth assignment. Data in column "Received" demonstrates the long-term fairness performance of P-WFQ discipline.

4.2 Delay Bound

The network topology used in simulations of this section is similar to that of Fig.2 in section IV.A. The differences are that the bandwidth between node 4 and 5 is changed to 5Mbps, time delay is 1ms, and queue buffer of node 4 is changed to 320 packets.

Table 1. Bandwidth Allocations of Different Algorithms

Displines	Node	Sent	Received	Percent
DropTail	0	25025	23774	0.95001
	1	25025	16	0.00064
	2	25025	16	0.00064
	3	25025	16	0.00064
WFQ	0	25000	11875	0.52252
	1	25000	5938	0.26124
	2	25000	2969	0.13064
	3	25000	2969	0.13064
WF^{2}Q+	0	24907	12375	0.49685
	1	25054	6438	0.25697
	2	24925	2989	0.11992
	3	24997	2988	0.11953
P − WFQ	0	25025	11895	0.47533
	1	25025	6135	0.24516
	2	25025	3122	0.12476
	3	25025	3121	0.12472

Table 2. Traffic Parameters of Each Nodes

Node	Flow Type	Rate(bps)	Packet Length(bytes)	Weight
0	CBR	64Kb	1000	1
1	CBR	1Mb	1000	2
2	Exponential①	2Mb	1000	2
3	FTP	--	1000	5

① burst-time and idle-time are 500ms respectively.

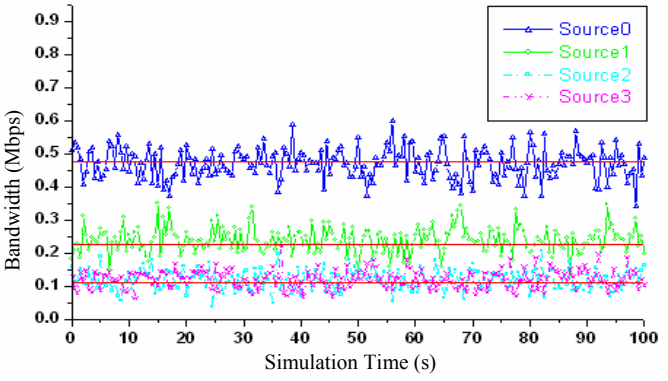


Fig. 3. Bandwidth Allocation – Fairness Examination

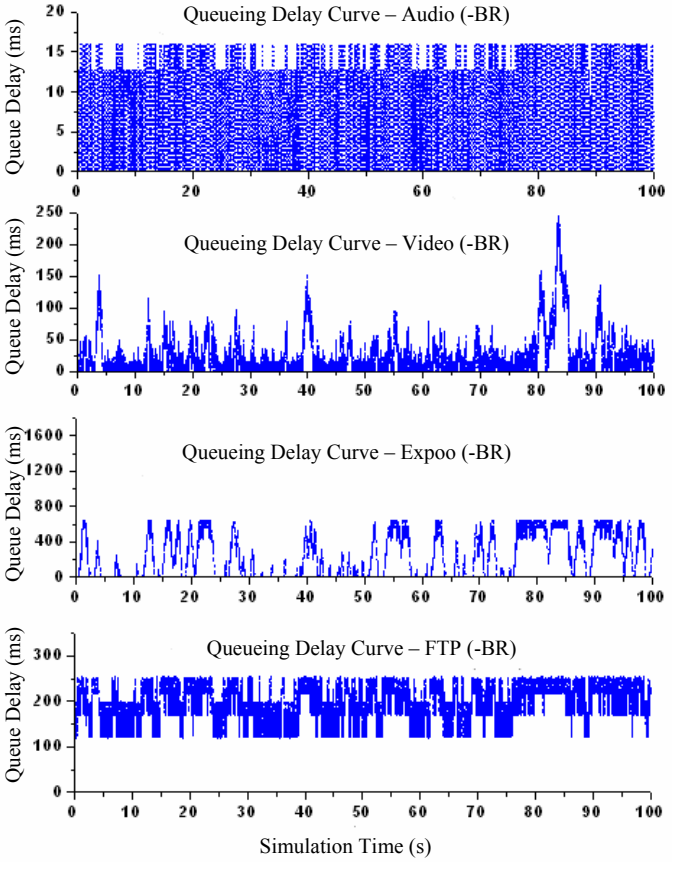


Fig. 4. Delay Curve – without BR

At the start, buffer length of each flow is 80 packets. The simulation time is 100s. Parameters of each flow are shown in table II. Data from node 0 represents audio data flow, node 1 video data flow, node 2 data flow with LRD (long range dependence), and node 3 the FTP data flow.

Fig.4 is the delay curves of each flow. Analysing Fig.4 we find that the audio and video data flows' queueing delays are limited within acceptable scopes. This shows that P-WFQ can grantee a queueing delay bound according to its weight. For time-sensitivity but loss-tolerant data flows, we also can limit queueing delay bounds by limit flows' buffer sizes. Then at the cost of few packets dropped, the fluctuation of queueing delay will be weakened.

5 Conclusion

In this paper we proposed a novel scheduling discipline for high-speed networks, referred to as Probability based Weighted Fair Queueing (P-WFQ) algorithm. It uses the serving probability parameter to implement the weighted server share. In this discipline, the tracing to change of weights of different flows commonly used in most weighted scheduling algorithms is avoided. This reduces the discipline's implementing complexity greatly. A new grouping algorithm is also proposed to satisfy with the requirement of high-speed networks. The simulation results demonstrate P-WFQ discipline's fairness and effectiveness for use in high-speed networks.

Acknowledgements

This work is partially supported by National Natural Science Foundation of China (No. 69984003) and National High-Technology Industrialization Project (No. 2050[2000]).

References

[1] A.Demers, S.Keshav, S.Shenker, Analysis and simulation of a fair queueing algorithm [J], *Internetworking: Research and Experience*, vol. 1, no. 1, pp. 3-26, 1990.

[2] M.Katevenis, S.Sidiropoulos, C. Courcoubetis, Weighted round-robin cell multiplexing in a general-purpose ATM switch chip [J], *IEEE J. Select. Areas Commun.*, vol. 9, pp. 1265–1279, Oct. 1991.

[3] A. K. Parekb ,R.G. Gallager. A generalized processor sharing approach to flow control in integrated services networks : the single-node case [J] .*IEEE/ ACM Trans. on Networking*, June 1993 :344 - 357.

[4] S.Golestani, A self-clocked fair queueing scheme for broadband applications [C], *Proc. IEEE INFOCOM'94* Toronto, Ont., Canada, pp. 636-646, Apr. 1994.

[5] M.Shreedhar, G.Varghese, Efficient fair queueing using deficit round robin [C], *Proc. ACM SIGCOMM'95*, Cambridge, MA, Sept. 1995, pp. 231–242.

[6] R.Bennett, H.Zhang. WF^2Q: Worst-case fair weighted fair Queueing [C]. *IEEE INFOCOM'96*, Mar. 1996 :120 - 128.

[7] P. Goyal, H. M. Vin, and H. Chen, “Start-time fair queueing: A scheduling algorithm for integrated services,” in *Proc. ACM-IGCOMM‘96*, Palo Alto, CA, pp. 157–168.

[8] J .C. R. Bennett, H. Zhang. Hierarchical packet fair Queueing algorithms [J]. *IEEE/ ACM Trans. on Networking*, Oct . 1997, 5 :675 - 689.

[9] O.Altintas ,et al . Urgency-based round robin : A new scheduling discipline for packet switching networks [C] . *IEEE INFOCOMM’98*, 1998 :1197 - 1183.

[10] Stephens, Donpaul C. ; Bennett, Jon C.R.; Zhang, Hui; "Implementing scheduling algorithms in high-speed networks[J]", *IEEE Journal on Selected Areas in Communications*, v 17, n 6, Jun, 1999, p 1145-1158

[11] Y.Jiang, C. K.Tham, C. C. Ko. A probabilistic priority scheduling discipline for multi-service networks [C]. *In IEEE ISCC’2001*, July 2001.

[12] Stoica, I.; Shenker, S.; Hui Zhang; Core-stateless fair queueing: a scalable architecture to approximate fair bandwidth allocations in high-speed networks [C]. Networking, IEEE/ACM Transactions onVolume 11, Issue 1, Feb. 2003 Page(s):33 - 46

[13] Network Simulator - ns-2 (2.28), http:// www.isi.edu/nsnam/ns/

Using of Intelligent Particle Swarm Optimization Algorithm to Synthesis the Index Modulation Profile of Narrow Ban Fiber Bragg Grating Filter

Yumin Liu[1,2] and Zhongyuan Yu[1,2]

[1] School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China
{Liu}liuyuminhqy@263.net
[2] Key Laboratory of Optical Communications and Lightwave Technologies Ministry of Education, Beijing 100876, China
{Yu}Yuzhongyuan30@hotmail.com

Abstract. A new method for synthesis of fiber Bragg gratings based filter is proposed. By combining the transmission matrix method and the particles swarm optimization algorithm, we obtain a novel method for the inverse problem of the synthesizing fiber gratings. With adjusting the parameters of the PSO algorithm we can get the demand index modulation for the target reflection spectrums including the phase response. Compared with other synthesis methods, the PSO algorithm characteristics are simple and faster convergence, especially by using the improved local PSO (LPSO) algorithm, we obtained the better results for the same problem.

1 Introduction

Fiber Bragg gratings have evolved into critical components for a multitude of applications in fiber communication systems [1]. The synthesis problems have attracted extensively interesting in the fields of physics and engineering. Recently lots of methods are proposed for the synthesis or reconstruction problems of the fiber Bragg grating [2]-[6]. It is important for both device design and characterization purpose.

The inverse Fourier transform of the grating spectrum can be accepted as the index modulation profile of the fiber Bragg grating for the weak grating especially for the maximum reflection peak is less 0.3 at which the first Born approximation is satisfied [2]. The Gel'fand-Levitan-Marchenko integral equations method is the exact description of the inverse problem of fiber Bragg gratings [3], but this method is very complex and is restricted to reflection spectrums which must be expressed as a rational function. Another powerful method is the layer-peeling techniques [4], which achieved successes for its simplex and easy to application, but it is very difficulty to design strong narrow band grating, otherwise the grating length required is typically long and the index modulation profile are complicated.

Evolution algorithm is a powerful artificial intelligence technology can be used for solving the problems of global optimization. Recently, fiber Bragg grating synthesis problems treated as a global optimization and solved by the evolution algorithm have been proved to be very useful [6][7][8].

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 438 – 447, 2006.

In this paper we present a new approach to the solution of inverse problems, in particular the problem of synthesizing fiber Bragg gratings. In our method the grating is divided into N sections, each section is treated as a uniform grating, so we can use the piece wise method to describe the whole grating and calculated the reflection spectrum. An error value is introduced by calculating the difference of the target spectrums and the calculated spectrums. Thus, the problem is a minimized problem of N variables. We use the particle swarm optimization (PSO) algorithm to solve the problem. The PSO was originated form the social modal and was presented by Kenedy and Eberhart in 1995. For its simplex and powerful, PSO algorithm have achieved great successes and become one of special topics by 'CEC'[8]-[13].

The remainder of the paper is organized as follows: First, in section II a short description of the principle of PSO is reviewed. Then in section III the grating theoretical approach of transmission matrix method was presented. Next, we discussed simulation results by three examples. Finally, the conclusion follows..

2 The Principle of PSO

The PSO algorithm can be applied to virtually any problem that can be expressed in terms of an objective function for which extremum must be found. The PSO algorithm is iterative and involves initializing a number of vectors (particles) randomly within the search space. The collective of the particles is known as the swarm. Each particle presents a potential solution to the problem of the target function. Each particle is also randomly initializing a vector called particle speed. During each time step the objective function is evaluated to establish the fitness of each particle using the particle itself as the input parameter. Then the particle will fly through the search space being attracted to both their personal best position as well as the best position found by the swarm so far.

We make the position of particle i expressed as $X_i = (x_{i1,}, x_{i2} \ldots, x_{id})$ and the best position of that particle so far expressed $p_i = (p_{i1}, p_{i2} \ldots, p_{id})$. The best position of the whole swarm can be expressed as $p_g = (p_{g1}, p_{g2}, \ldots, p_{gd})$. Then the particle position update can be expressed as $x_{id} = x_{id} + v_{id}$ where the v_{id} denote the speed of the d dimension component of particle i and expressed as:

$$v_{id} = wv_{id} + \phi_1 rand()(p_{id} - x_{id}) + \phi_2 rand()(p_{gd} - x_{gd}) \tag{1}$$

where w is the inertia weight determining how much of the particle's previous speed is preserved, ϕ_1, ϕ_2 are two acceleration constants present the cognition part and the social part respectively, $rand()$ is uniform random sequences from {0,1}. In the stand PSO algorithm (1) only the inertia weight is introduced. To improve the performance, we introduced another parameter K expressed as: $K = 2 \cdot \left| 2 - \phi - \sqrt{\phi^2 - 4\phi} \right|^{-1}$, $\varphi = \varphi_1 + \varphi_2$, $\varphi > 4$, so the speed update can be expressed as:

$$v_{id} = K\left(wv_{id} + \varphi_1 rand()(p_{id} - x_{id}) + \varphi_2 rand()\left(p_{gd} - x_{id}\right)\right) \tag{2}$$

The iterative process will continue using the formula (2) until the extremum has been found or the number of iteration access to the maximum value. Fig.1 shows the flow chart of the particular PSO algorithm that we used for the synthesis of the fiber Bragg gratings.

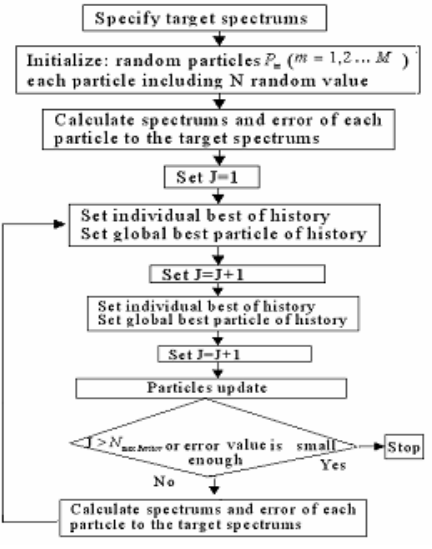


Fig. 1. Flow chart of the PSO algorithm for the synthesis of fiber Bragg gratings

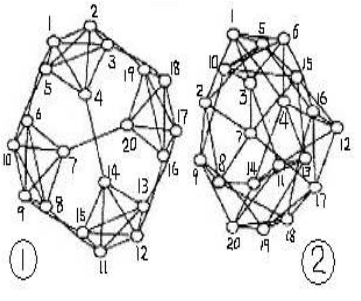


Fig. 2. The two possible topologies of the swarm particles for the LPSO

Generally, there are two kinds of topological neighborhood structures, global neighborhood structure, corresponding to the global version of PSO (GPSO), and local neighborhood structure, corresponding to the local version of the PSO (LPSO). For the global neighborhood structure the whole swarm is considered as the neighborhood, while for the local neighborhood structure some smaller number of adjacent members in sub-swarm is taken as the neighborhood. In the GPSO, each particle's search is influenced by the best position found by any member of the entire population. In contrast, in LPSO, the search is influenced only by parts of the adjacent members. It is

widely believed that GPSO converges quickly to an optimum but has the weakness of being trapped in local optima occasionally, while the LPSO is able to "flow around" local optima, because the sup-swarm explore different regions in hyperspace. To improve the search ability, for the simple examples A and B in section IV, we used the GPSO algorithm. For the complex example C, we used the LPSO algorithm, and the two possible topologies for LPSO are shown in Fig.2 [14].

3 The Principle of Fiber Bragg Gratings

The index modulation of the FBG is a periodic function, and can be expressed as cosine function:

$$n(z) = n_0 + \Delta n(z)\cos(2\pi z / \Lambda + \theta(z)) . \tag{3}$$

where $\Delta n(z), \theta(z)$ are the amplitude and phase modulation of the fiber Bragg grating respectively, and slowly varying compared to the grating period. There are two contra direction couple mode in single mode fiber Bragg grating and expressed as:

$$\begin{aligned} dA/dz &= -ikB - i\delta A \\ dB/dz &= -ik^* A + i\delta B \end{aligned} \tag{4}$$

where A and B are the amplitude envelopes of the two modes , k is the couple efficient, δ is the detuning $\delta = \pi\Delta n / \lambda - \pi / \Lambda$. For the un-uniform fiber Bragg grating one cannot get the analytic expression. To analysis the transmission characteristics of such grating, one must introduce numerical solution. A lot of methods can be used to describe the differential equation of (4), such as Runge-Kutta algorithm, matrix transmission method, predictor-corrector and finite difference method. In this paper we use the matrix transmission method, where the grating is uniformly divided into N sections and each section is treat as a uniform grating expressed as a matrix of (5). Then the whole grating can be expressed as multiplication of each matrix:

$$T_i = \begin{bmatrix} \cosh(\Omega_i dz) - i\delta_i \sinh(\Omega_i dz)/\Omega_i & -ik\sin(\Omega_i dz)/\Omega_i \\ ik\sin(\Omega_i dz)/\Omega_i & \cosh(\Omega_i dz) + i\delta_i \sinh(\Omega_i dz)/\Omega_i \end{bmatrix} \tag{5}$$

$$T = T_m \cdot \ldots T_{m-1} T_k T_1 \tag{6}$$

where $\Omega_i = \sqrt{k_i{}^2 - \delta_i{}^2}$, and the complex reflective amplitude of the grating is $r = T_{11}/T_{21}$. In this paper we set the error function as:

$$\mathrm{Err}\{R_{calc}, R_{target}\} = \frac{1}{J}\sum_j (R_{target,j} - R_{clac,j})^2 \tag{7}$$

where $R_{target,j}$ and $R_{calc,j}$ are the target and calculated spectrum at the samples of j position of the reflection spectrum, and $R_{calc,j} = r_j^2$. As shown in the flow chart of Fig.1, we can use the following steps to optimize the target spectrum by the PSO algorithm: Randomly Initialize M particles, each particle is a D dimension random numbers as the index modulation profile, D is the number of section of the grating,

and D is in real space, M is the number of swarm. Compare the error value of each particle with the minimum error value of its history. If the current error value is smaller than the history error value of the particle, the history position and the error value of the particle are replaced by the current particle position and the error value. Otherwise nothing changed. Updating the position and speed of each particle for the next loop until the optimization condition is satisfied or the maximum iteration number is reached.

4 Numerical Results and Discussions

4.1 Bandpass Filter

First, we use the synthesis algorithm to synthesize a fiber band pass filter. The target band pass filter is characterized using an ideal rectangle band pass filter.

$$R = r^2 = \begin{cases} 1 & |\delta| < a \\ 0 & |\delta| > a \end{cases} \tag{8}$$

where $2a$ refers to the band width of the filter. The pass band is reflection band, and in this case we set the band to be $0.2nm$. The central Bragg grating is $1550nm$, the maximum index modulation is 5×10^{-4}, the length of the grating is $1cm$, the grating is divided into 16 sections, the particles is generated randomly, and the swarm is composed of 250 particles because the dimension of the particle is too large. The best index modulation profile optimized after 20 iterations by the PSO is shown in Fig.3 (a), and the reflectivity spectrum is shown in the solid curve of Fig.3(b). The best index modulation is just like a raise-cosine function. The side band is suppressed approximately to be less than $-25db$. Fig.4(a) and Fig.4 (b) gave the best index modulation and reflectivity spectrum after 100 iterations calculation by the GPSO algorithm. We can see from Fig.4 (b) that the capability of the side band suppression

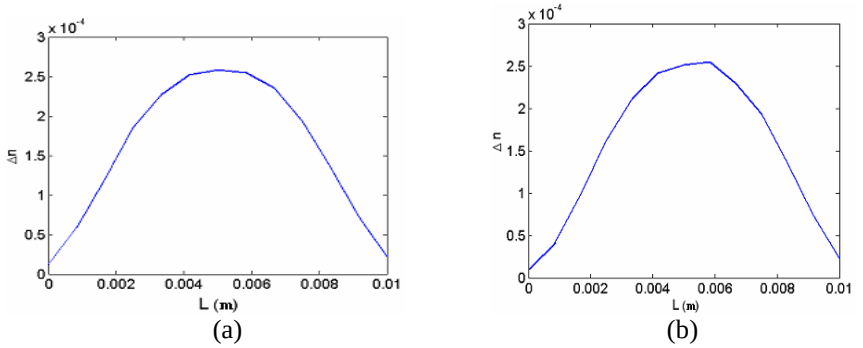


Fig. 3. The calculated index modulation profile of the best particles in the 20 generation (a), and 100 generation (b)

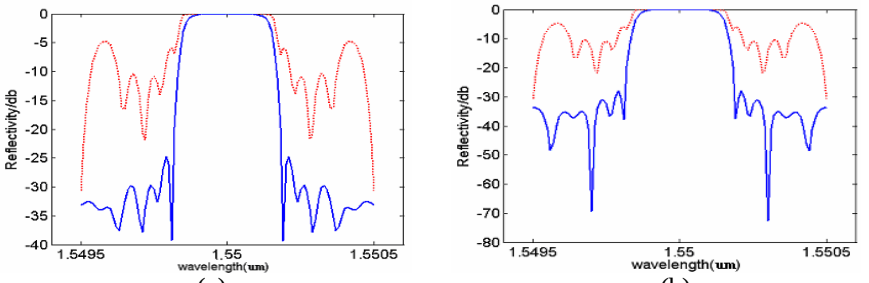


Fig. 4. Calculated reflection spectrums. (a)The dotter curve is the best spectrum of the fist generation; the solid curve is the best spectrum of 20 generation; (b)The dotter curve is the best spectrum of the fist generation; the solid curve is the best spectrum of 100 generation.

is further improved, and the side band is less than $-30db$. As a comparing with the article [6], where the genetic algorithm solves the inverse problem of band pass filter, the PSO algorithm is more efficient and the speed of convergence is faster. The reflection spectrum is accepted only after 10 iterations. We found that after 100 iterations the index modulation convergence speed is very slow and we can assume that it has reached the best position. The optimized results in our algorithm are better than that in literature [6]. The first side band of the grating synthesized by the GPSO is less –30dB, while in the genetic method the first band is only about –20dB. Real-coded genetic algorithm and simulated annealing algorithm have been proved their usefulness for fiber Bragg grating parameters synthesis [8,17], however there are only three parameters to be synthesized, so we thought it is a trivial problem and any optimized method (including PSO algorithm off course) could fulfill such tasks.

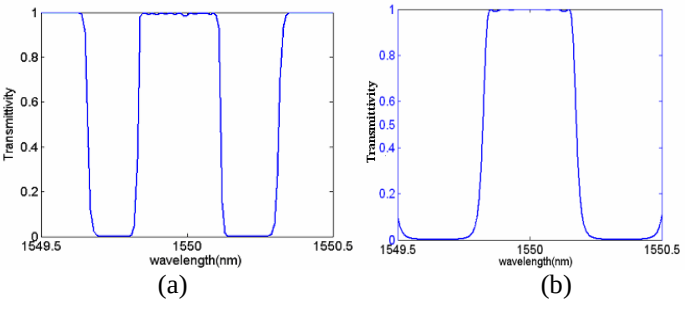


Fig. 5. The optimized transmittivity spectrums of the multi phase-shift grating; (a)when the phase-shift positions is uniformly distributed along the grating, (b) when the positions of phase-shifts along the grating are optimized at the fixed phase-shift values listed at Table.1

4.2 Multi Phase-Shift Transmission Filter Improved by the PSO

The fiber Bragg grating filter can be designed as a transmission filter by introducing multi phase-shift along the grating. But the positions and amplitude of the phase-shift,

which determines the transmission characteristics, is a critical problem. In the flowing we first optimize the amplitudes of the phase-shift, and based on which we optimize the positions of the multi phase-shift. We adopt the index modulation profile of the Fig.3 (b) as the multiphase-shift grating index amplitude modulation profile and the target of the transmittivity spectrums is expressed as:

$$T = t^2 = \begin{cases} 1 & |\delta| < a \\ 0 & |\delta| > a \end{cases} \tag{9}$$

where a is 0.2nm. In the first step we fix the positions of phase-shift along the grating uniformly. The amplitudes of each phase-shift are the objects we want to optimize. In the next step we fix the optimized phase-shift and optimized the positions of the phase-shift along the grating. The number of the phase-shift introduced in the grating is assumed to be 11. Amplitude of the phase-shift optimized by our PSO algorithm is listed in table 1, and the transmittivity spectrums are shown in Fig.5 (a)

Table 1. Amplitude of phase-shift along the grating

Phase1	2.8032	Phase7	2.8765
Phase2	-2.8386	Phase8	3.3092
Phase3	2.8860	Phase9	2.8574
Phase4	2.8975	Phase10	2.8250
Phase5	2.9003	Phase11	-9.7444
Phase6	2.8906		

Table 2. The position of phase-shift along the grating (mm)

Z1	0.0594	Z7	0.5842
Z2	0.1386	Z8	0.7129
Z3	0.1782	Z9	0.8416
Z4	0.2376	Z10	0.8712
Z5	0.3465	Z11	0.9703
Z6	0.4653		

Using the date of table1, we continue the next optimization process where the amplitudes the phase-shift is fixed just as the data listed in table 1, and the positions along the grating is the objects that should be optimized. The optimized dates are listed in Table.2. The optimized transmittivity spectrums are shown in Fig.5 (b).

Comparing with the Fig.5 (a) and the Fig.5 (b), we can see the transmittivity spectrums of Fig.5 (b) are more approach the target spectrums of (9) than that of Fig.5 (a). It is showed that the second optimization is necessary when only the amplitudes of phase-shift optimization cannot acquire the demand of target spectrum. Another noticeable problem is that the optimized parameters of the problems of synthesis fiber Bragg grating must be reasonable. In other words, the optimized range of the parameters must reasonable to realize the target spectrum to be optimized, which must be judge by experience.

4.3 Low Dispersion Narrow-Band Filter

Synthesizing a low-dispersion narrow-band filer using the fiber Bragg grating by PSO is a complicated multi-objective optimization problem. To get better results we used the LPSO algorithm, and the topological structures are just as shown in Fig.2. The error function of the particle is the weighing of the amplitude and phase response and expressed as:

$$\text{Err}\{R_{cal}, R_{target}, D_{cal}, D_{target}\} = (1-b)\frac{1}{J}\sum_{j}\left(\frac{R_{target,j}-R_{cal,j}}{\sigma_j}\right)+b\frac{1}{K}\sum_{k(inband)}(D_{target,k}-D_{cal,k}) \quad (10)$$

where $D_{target,k}$ and $D_{calc,k}$ are the target and calculated dispersion at the sampling point k, and we only consider the dispersion response in reflective band, b is the weighting factor between the reflectivity and dispersion, in our simulation, we assumed $\sigma_j = 0.001$, the unit of dispersion is ps/nm, $b = 0.5$, and the grating length is 4.5cm. To prove the ability of LPSO, in designing the dispersion-free filter, we divided the grating into 80 sections. This can assure the smooth index modulation for designing complex filter. Unlike dividing the grating into 20 sections [7], which make the optimized the index modulation is an obviously stepped function and difficult to fabrication.

There are 20 particles in each swarm, and five swarms are used in our simulation. Fig. 6 (a) and (b) shows the optimized amplitude and phase response using the LPSO algorithm after 500 generations, The first side band is less than –25dB,and the in-band delay ripple is less than ± 3 ps. Fig. 6 (c) shows the optimized index modulation profile, as can be seen a smooth index modulation can be obtained by our LPSO algorithm.

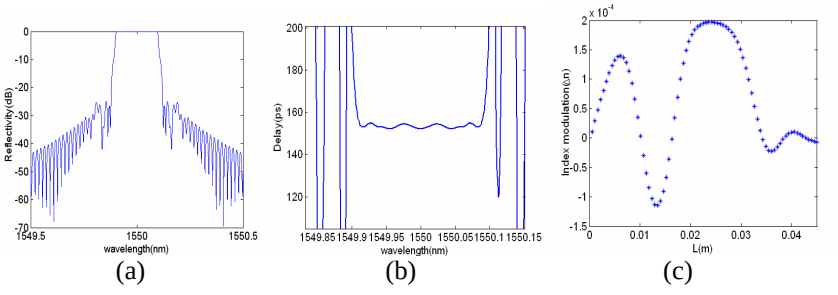
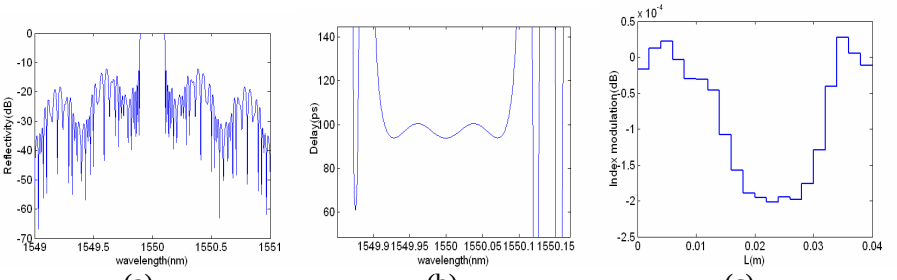


Fig. 6. The optimized reflection spectrum (a), delay spectrum (b) and best index modulation by the LPSO algorithm after 500 generation

As a comparison with the results obtained by genetic optimization [6], we found that LPSO algorithm can provide better performance in designing narrow-band low dispersion filter, including both amplitude and phase response. In Johannes Skaar's literature [6], only the phase can be acceptable, the first side band of the amplitude response is large than –10dB. In fact, such bad performance filters cannot be used for optical communications system. For comparison purpose, in Fig.7, we gave the optimized results by an improved version of the genetic algorithm, which is a standard toolbox of the Matlab software. The err function and other physical parameters are just the same with that used in the LPSO. In the improved version we used the sub swarm technology. As can be seen, the best results after 1000 generations are better

than that of the literature [6]. As the knowledge about author, although genetic algorithm has been used for synthesis fiber Bragg grating or obtained some applications in strain sensor [15-16], however, better results can only be obtained when the applications are dependent only on the amplitude response. For applicators that dependent on both amplitude and phase response, no acceptable results have been reported yet.



(a) (b) (c)

Fig. 7. The optimized reflection spectrum (a), delay spectrum (b) and best index modulation by the genetic evolution algorithm after 1000 generation

5 Conclusions

In this paper we present a new method to solve the problems of synthesis fiber Bragg gratings. Using three design examples: band pass filter, multi-phase band pass filter and narrowband dispersion-free filter, the ability of the PSO, especially the version of LPSO was proved to be powerful and efficient. Compared with the other algorithms the advantages of the PSO are that its simplicity and the convergence speed is very fast so reduced calculating time-consume greatly. We compared the optimized results of the complex narrowband dispersion-free filter with those obtained by the genetic algorithm and Layer peeling algorithm. We found that the results obtained by LPSO are better than these methods both in the aspects of amplitude and phase response. Otherwise, the LPSO method can also be used for other synthesis problem, especially for photoelectronic devices [14].

Acknowledgment

This work was supported by the National “973” Basic Research and Development Program of China (No.2003CB314901), the National “863” High Technology Project of China (No. No.2003AA311070).

References

1. Morten Ibsen and Ricardo Feced, “Fiber Bragg gratings for pure dispersion-slope compensation,” *Opt.Lett.*, vol.28, pp.980-982,2003.
2. L.Poladian, “Simple grating synthesis algorithm,” *Opt.Lett.*, vol.25, pp.787-789,2000.

3. L.Wang and T.Erdogan, "Layer Peeling algorithm for reconstruction of long-period fiber gratings," *Electron.Lett*.,vol.37, pp.154-155,2001.
4. Jan Kristoffer Brenne and Johannes Skaar, "Design of grating-assisted codirectional couplers with discrete inverse-scattering algorithms," *J.Lightwave Technol*., vol.21, pp.254-263,2003.
5. Eva Peral, Jose Capmany and JavierMarti, "Iterative solution to the gel'fand-levitan-marchenko coupled equations and application to synthesis of fiber gratings," *IEEE J.Quantum Electron*, vol.32, pp.2078-2084,1996.
6. Johammes Skaar and Knut Magne Risvik, "A genetic algorithm for the inverse problem in synthesis of fiber gratings," *J.Lithtwave Technol*, vol.16, pp.1928-1932, 1998.
7. C.L.Lee and Y.Lai, "Evolutionary programming synthesis of optimal long-period fiber gratings filters for EDFA Gain Flattening," *IEEE Photon. Technol. Lett.*, vol.14, pp.1557-1559,2002.
8. Po Dong, Jose Azana, and Andrew G.Kirk, "Opt. Communications, 228 (2003) 303
9. K. E. Parsopoulos and M. N. Vrahatis, "Particle swarm optimizer in noisy and continuously changing environments", *Artifical Intelligence and Soft Computing*, IASTED/ACTA Press, pp289-294, 2001.
10. J. Kennedy and R. C. Eberhart, "Particle swarm optimization ", *Proc. Of IEEE International Conference on Neural Networks*, Piscataway, NJ, USA, pp1942-1948, 1995.
11. Y. Shi and R. C. Eberhart, "Parameter selection in particle swarm optimization", *Evolutionary Programming VII*, pp591-600, 1998
12. M. Clerc and J. Kennedy, "The particle swarm-explosion, stability, and convergence in a multidimensional complex space", *IEEE Trans. on Evolutionary Computation*, v6, pp58-73, 2002,
13. R. C. Eberhart and Y. Shi, "Comparison Inertia weights and constriction factors in particle swam optimization ", *Proc. 2000, ICEC*, pp84-88, 2000
14. Yumin Liu, Zhongyuan Yu, Hongbo Yang, Na Zhang, Qiang Feng, Xiaogguang Zhang, "Numerical optimization and simulation to wavelength-division multiplexing isolation filter consisted of two identical long period fiber grating" *Optics Communications*,246, pp367-372, 2005.
15. Cinzia Zuffada, Tom Cwik, Christopher Ditchman, "Synthesis of Novel All-Dielectric gratings filter using Genetic Algorithms", IEEE Transctions on Antennas and Propagation, Vol.46 pp657-663, 1998.
16. Federico Casagrande, Paola Crespi, Anna Maria Grassi *etl.* "From the reflected spectrum to the properties of a fiber Bragg grating: a genetic algorithm approach with application to distributed strain sensing", Applied optics, Vol.41, pp5238-5343, 2002.
17. Gabril Cormier, Roger Boudreau, "Real-coded genetic algorithm for Bragg grating parameter synthesis", Optical Society of America B, Vol.18, pp1771-1776, 2001.

Chaotically Masking Traffic Pattern to Prevent Traffic Pattern Analysis Attacks for Mission Critical Applications in Computer Communication Networks

Ming Li[1] and Huamin Feng[2]

[1] School of Information Science & Technology, East China Normal University, Shanghai 200062, PR. China
mli@ee.ecnu.edu.cn, ming_lihk@yahoo.com
[2] Key Laboratory of Security and Secrecy of Information, Beijing Electronic Science and Technology Institute, Beijing 100070, PR. China
fenghm@besti.edu.cn

Abstract. We propose a novel approach to chaotically mask traffic pattern according to a predetermined chaotic mode, aiming at making the masked traffic lose values for intruders to analyze for malicious activities in a mission critical network.

1 Introduction

Information security in the Internet is expected to be guaranteed by security protocols, such as IP security protocol [1] and transport layer security [2]. However, it is not enough for a mission critical application, e.g., military networks, just to encrypt and authenticate transmitted data since a real-traffic series has its statistical patterns [3], which contain information for intruders to analyze to launch so called traffic analysis attacks [4]. Thus, methods to prevent traffic analysis attacks are desired in ultra-secure distributed systems. The idea behind that is based on Shannon's perfect secrecy theory as stated like this. If one can map any masked traffic to a predefined pattern, which is a sufficient condition used by most researchers, the adversary cannot obtain any information by analyzing the masked traffic [5].

Several approaches have been reported, see e.g. [6], [7]. The work [8] gave a scheme to mask traffic pattern using constant packet size. Though simple in method, it has the disadvantage since opponents may easily find secure network if traffic with constant packet size is identified. This paper substantially extends [8] so that traffic pattern analysis is prevented based on chaotically masking traffic pattern on a link-by-link basis. Besides, we also hide IP addresses of each packet so as to enhance difficulties for intruders to trace the packet within a time duration they desire.

This paper is organized as follows. Section 2 discusses the methodology. A case study is given in Section 3 and conclusions in Section 4.

2 Methodology

We purposely control the lengths of the packets to be sent by one network node based on a predetermined chaotic mode with IP addresses of packets encrypting. During

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 448 – 451, 2006.

each unit time period, only one packet is sent. If the length of a packet to be sent is shorter than expected length, some extra bytes are padded to its rear. If longer than expected length, the packet is divided into two or more. The length of each packet is aligned according to a chaotic data generator that obeys a predetermined chaotic mode. Dummy packet and extra bytes make up the complementary traffic.

In the present scheme, masking traffic pattern in overall network is done in the same way as that in each network segment. Hence, it is on a link-by-link basis. A side benefit of doing so is that the cost of secure communications in the complementary traffic does not consume the bandwidth of other network segment though such a secure strategy is generally under the assumption that bandwidth is always sufficient.

The above scheme contains 3 components: 1) a chaotic data series generator; 2) an aligner algorithm to align packet length according to a given chaotic mode; and 3) a protocol to negotiate the parameters, such as the encryption algorithm and key. Since 3) is similar to that in [8], the following will only discuss 1) and 2).

Denote a network by a directed graph (N, E), where $n \in N$ is a network node (e.g., host or router). Then, for $n_1, n_2 \in N$, if $(n_1, n_2) \in E$, (n_1, n_2) is a link e. Denote len_e the maximum length (bytes) of a frame in e and σ_e as the minimum interval time between two frames. Define two terms as follows. The first one is link feature and the second man-made traffic pattern. The feature of link $e \in E$ means a pair (len_e, σ_e). The man-made traffic pattern implies that $x_e[t(i)]$ for e has a predetermined chaotic mode $c_e(x)$, if $x_e[t(i)]$ follows $c_e(x)$ and $0 \le x_e[t(i)] \le len_e \ \forall e \in E$, where $x_e[t(i)]$ is the length of the ith packet for e at the time $t(i)$, where $t(i)$ is the timestamp ($i = 0, 1, 2, \ldots$).

Chaotic data generator is described as follows. Solutions to non-linear differential equations are dependent on initial conditions [9]. A solution of a non-linear system in time is a time series. We note that a significant advantage using chaotic pattern is that chaotic series in theory is unpredictable. The following uses the simple pendulum just for the sake of indicating a case of chaotic data generation.

Let $x(t)$ be the angle displacement of the pendulum and m be its mass. Let l be the length of string and g be the gravitation acceleration (9.8 m/s^2). Suppose a periodic force exerts on the pendulum such that the Lagrangian function $L(x, \dot{x})$ of the pendulum system, without considering the damping, is

$$L(x, \dot{x}) = \frac{1}{2} ml^2 \dot{x}^2 + mgl\cos(x) - k\sin(\omega t)\dot{x}, \tag{1}$$

where k is a constant. The Lagrangian equation of motion with respect to (1) is given by

$$\ddot{x} = -\frac{g}{l}\sin(x) + \frac{k}{ml^2}\omega\cos(\omega t). \tag{2}$$

Let $-B\dot{x}$ be the damping. Then, considering the damping, the system is given by

$$\ddot{x} = -\frac{g}{l}\sin(x) + A\cos(\omega t) - B\dot{x}, \tag{3}$$

where $A = (k/ml^2)\omega$. Let $\omega = 2/3$, $l = g$, $A = 1.5$ and $B = 0.5$. Then, Fig. 1 and Fig. 2 show a phase plot and its corresponding chaotic time series at $x(0) = -2$, $x'(0) = 3$.

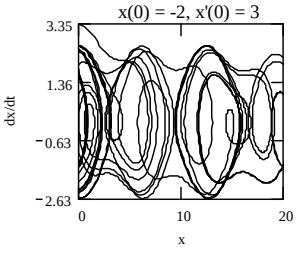


Fig. 1. Phase plot at $x(0) = -2$, $x'(0) = 3$

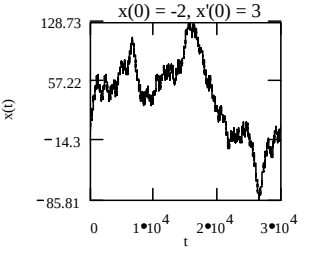


Fig. 2. Chaotic series at $x(0) = -2$, $x'(0) = 3$

The functionality of the aligner is stated as follows. For a given traffic series $x[t(i)]$, the aligner should transform $x[t(i)]$ into $y[t(i)]$ such that y obeys (3) under given initial conditions. In passing, encryption operation needs keys and algorithms. Following the IKE protocol [10], we define the link security association of $e \in E$ by a pair $sa_e = (Enc_e, k_e)$, where Enc_e is the encryption algorithm while k_e implies the relevant key. The addresses of packet transmitting on e can be encrypted by Enc_e with k_e.

Man-made traffic on e can be detailed like this. During the interval σ_e, if there is no packet to be transmitted on e, we need to send a dummy packet with len_e, or if the size of packet transmitting on e is less than len_e, some extra bytes are appended at the packet rear so that it has the size len_e.

3 A Case Study

Fig. 3 (a) indicates a time series of real traffic named pAug.TL [11]. Fig. 3 (b) is its autocorrelation function (ACF for short). Let $x(t)$ in Fig. 3 (a) be the traffic to be masked. Calibrate the series in Fig. 2 to be in the range of 64 to 1518 bytes. Aligning $x(t)$ with that calibrated chaotic series yields a chaotically masked series $y(t)$ as shown in Fig. 4. Let $R_y(k)$ be the ACF of $y(t)$. Fig. 5 indicates $R_y(k)$. Let corr$[R_y, R_x]$ be the correlation coefficient between R_y and R_x. By computation, we obtain corr$[R_y, R_x] = 0.21$, implying that R_y significantly differs from R_s. Hence, the pattern of x is hidden.

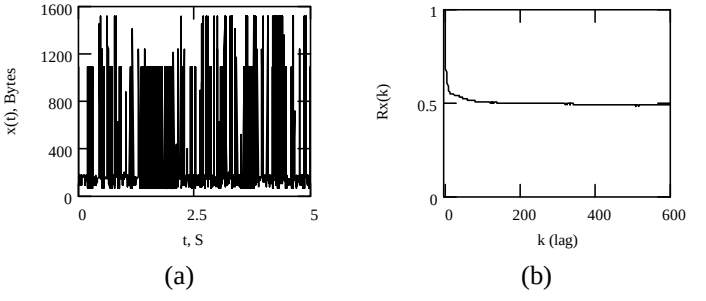


Fig. 3. A real traffic series and its autocorrelation. (a) Time series x. (b) Autocorrelation of x.

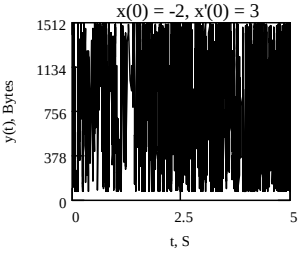


Fig. 4. Masked traffic

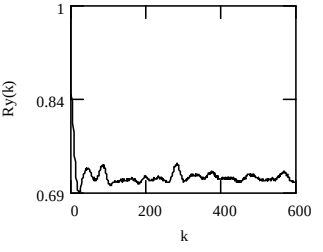


Fig. 5. Autocorrelation of masked traffic

4 Conclusions

We have explained an approach to prevent traffic pattern analysis. An advantage and a reason to use chaotically masking is because a chaotic series is sensitive to initial conditions so that one may easily obtain flexible data sets for aligning masked traffic.

Acknowledgements

Anonymous referees' comments on this paper are appreciated. This work was partly supported by the National Natural Science foundation of China under the project grant number 60573125.

References

1. Kent, S., Atkinson, R.: Security Architecture for the Internet Protocol. RFC 2401
2. Dierks, T., Allen, C.: The TLS Protocol Version 1.0. RFC 2246
3. Li, M.: Computer & Security 23 (2004) 549-558; 25 (2006) 213-220
4. Baran, P.: On Distributed Communications: IX Security, Secrecy, and Tamper-Free Considerations. Memo RM-3765-PR, Rand Corp. (1964)
5. Shannon, C. E.: Communication Theory of Secrecy Systems. Bell Sys. Tech. J. 28 (1949) 656–715
6. Deng, J., Han, R., Mishra, S.: Intrusion Tolerance and Anti-Traffic Analysis Strategies for Wireless Sensor Networks. Int. Conf., Dependable Systems and Networks (2004) 594-603
7. Deng, J., Han, R., Mishra, S.: Countermeasures Against Traffic Analysis Attacks in Wireless Sensor Networks. TR CU-CS-987-04, University of Colorado at Boulder (2004)
8. Liu, D. X., Chi, C.-H., Li, M.: Normalizing Traffic Pattern with Anonymity for Mission Critical Applications. 37th Annu. Simulation Symposium, Virginia, (2004) 293-299
9. Alligood, K. T., Sauer, T. D., Yorke, J. A.: Chaos: An Introduction to Dynamical Systems. Springer (1997)
10. Harkins, D., Carrel, D.: The Internet Key Exchange (IKE). RFC 2409
11. http://www.acm.org/sigs/sigcomm/ITA/

A New Secure Communication Scheme Based on Synchronization of Chaotic System

Yonghong Chen

Dept. of Computer & mathematics, Chongqing Normal University, 400047, China
djandcyh@163.com

Abstract. It proposes a chaotic secure communication scheme by introducing a concept of key function (KF) and key initial conditions (KICs). By using KF, secret signal is combined with complex chaotic signal in order to enhance the security of transmitted signal. KICs are used to enlarge minor mismatch of drive signal so as to increase initial conditions sensitivity of the auxiliary driver system. Experiment results show that proposed cryptosystems are very sensitive to KICs' and KFs' mismatch.

1 Introduction

Recently, chaotic system and its useful application in secure communication have been widely studied. Since Pecora and Carroll [1] proposed chaos synchronization and verified by electric later, chaotic circuits and applications for secure communications have received a great deal of attention [2],[3]. However, it have shown that most of them are not equipped with sufficient security [4],[5],[6].

In this paper, encryption method is based on EKF to generate a sufficiently complicated signal, which is compounded by masked secret information and chaotic signal. By this method, we make use of an auxiliary chaotic system embedded in the encrypter and the decrypter to drive two identical chaotic systems to synchronize. During decryption processing, secret information can be retrieved correctly if and only if decrypter offers the same KICs and right DKF corresponding to EKF. This design mechanism can actually make communication system more secure.

2 Chaotic Secure Communication Systems

In this section, it will present chaotic secure communication systems for transmit signals. System block diagram is shown in Fig. 1. Secure system has two major parts: encrypter and decrypter.

Encrypter consists of an auxiliary driver chaotic system, a sender chaotic system and EKF combination system. Decrypter is composed of an identical receiver chaotic system, an auxiliary driver chaotic system and a corresponding DKF decompound system. Signal of auxiliary system is to synchronize two identical systems embedded in encrypter and decrypter. In following text it will be described.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 452 – 455, 2006.

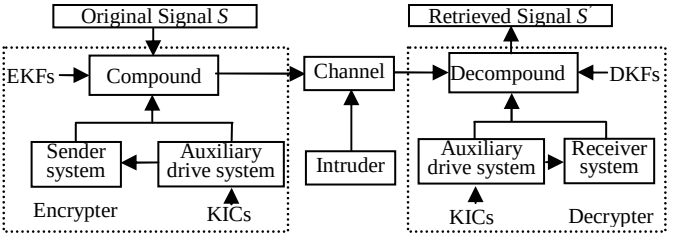


Fig. 1. Chaotic secure communication system block diagram

2.1 Synchronization Scheme

Assume auxiliary driver chaotic system, sender chaotic system and receiver chaotic system are respectively defined as following forms

$$\{X'=f(X_1, X_2),\ Y'=g(X_1, Y)\ ,\ Z'=g(X_1, Z)\ \}^T . \tag{1}$$

where $X_1=(x_1, x_2,\ldots, x_l\ )^T$, $X_2=(x_{l+1}, x_{l+2},\ldots, x_n\ )^T$, $Y=(y_1, y_2,\ldots, y_m\ )^T$ and $Z=(z_1, z_2,\ldots, z_m\ )^T$, $f()$ and $g()$ are nonlinear vector valued functions. First part of Eq. (1) is called auxiliary drive system which generate drive signal X_1, while second and third part of Eqs. (1) are called sender and receiver systems which will be synchronized by driver signal X_1. Synchronization is said to be achieved if get

$$e(t)= Z(t)\text{-}Y(t)\rightarrow 0 \qquad \text{as } t\rightarrow\infty . \tag{2}$$

An effective synchronized to second part method is to reconstruct an identical third part to second part, and driving it with same X_1. But if different initial conditions to start system (1), it will generate different X_1, then second and third part cannot be synchronized. Rössler system is well-known chaotic system, define as:

$$\{x'_1=\text{-}x_2\text{-}x_3,\ x'_2=x_1+0.2x_2,\ x'_3=0.2+x_3(x_1\text{-}5.7)\}^T . \tag{3}$$

Use its output $x_1(t)$ as drive signal to drive two identical unified chaotic systems [6]

$$\{y'_1=(25\rho+10)(\ y_2\text{-}\ y_1),\ y'_2=(28\text{-}35\rho)x_1\text{-}x_1\ y_3\text{-}(29\rho\text{-}1)\ y_2,\ y'_3=\ x_1\ y_2\text{-}(\rho+8)/3y_3\ \}^T. \tag{4}$$

Eq.(4) is a sender system, receiver systems is defined as

$$\{z'_1=(25\rho+10)(\ z_2\text{-}\ z_1),\ z'_2=(28\text{-}35\rho)x_1\text{-}\ x_1\ z_3\text{-}\ (29\rho\text{-}1)\ z_2\ ,\ z'_3=\ x_1z_2\text{-}(\rho+8)/3\ z_3\}^T . \tag{5}$$

In [6], sufficient conditions for chaos synchronization of the unified chaotic system are gained. If $0\leq\rho<1/29$, system (5) will synchronize to system (4).

2.2 Secure Communication for Transmission Signals

Based on synchronization method in section Fig.1, we proposed a secure communication scheme for transmission signals in the following text. In Fig.1. Chaotic secure communication system, the transmitted signal $T(t)$ of encrypter is defined as

$$T(t)=AX_1+\alpha g(X_2,S(t)) . \tag{6}$$

where A is a constant matrix with appropriate dimension, $\alpha(\neq 0)$ is a scaling factor allowing the term $g(.)$ to belong to a compatible range with respect to the minimum

and maximum bound of capability. S is the external input secret information, which is to be masked to transmit. $g(X_2,S(t))$ is an appropriate EKF, maybe linear or nonlinear which is chosen by both sides encrypter and decrypter. The reason is that it is to generate a sufficiently complicated chaotic signal, which consists of the secret information S to be masked. Form of EKF is within our choice and possible choice functions are massive. So S is combined to generate transmitted signal T by EKF.

Assume received signal from channel is T. Then retrieved secret information is defined as

$$S'(t)=h[X_2',1/\alpha(T(t)- AX_1)] \ . \tag{7}$$

$h[X_2',1/\alpha(T(t)- AX_1)]$ is a DKF corresponding to EKF $g(X_2,S(t))$.

3 Experimental Results

In this experiment, system (3), (4), (5) are auxiliary driver systems, sender and receiver system. Set KICs is $(0.1, 0.2, 0.3)^T$, EKF is $(y_1+ y_2+ y_3)+(S(t)+2)(x_2^2+x_3^2+50)$. DKF is $(T(t)-z_1-z_2- z_3)(x_2^2+x_3^2+50)-2$. Original information $S(t)=0.5\sin(0.1t)$ is secure signal, it is shown in Fig.2(a). When auxiliary driver systems, sender and receiver system are starting from $(0.1,0.2,0.3)^T$, $(0.1,0.1,0.1)^T$, $(1,1,1)^T$, the transmitted compound signal $T(t)$ versus t is shown in Fig.2(b). The correctly retrieved signal $S'(t)$ is shown in Fig.2(c). The error $e(t) =S'(t)-S(t)$ is shown in Fig.2(d). When auxiliary driver systems are starting from little different initial conditions $(0.1,0.2,0.3)^T$ and $(0.1001,0.2,0.3)^T$, sender and receiver system are starting from the same initial conditions $(0.1,0.1,0.1)^T$, EKF and DKF are the same as above. The numerical simulations results of incorrectly retrieved signal are shown in Fig.3(a). The error $e(t)$ is shown in Fig.3(b). When auxiliary driver systems, sender and receiver system are starting from $(0.1, 0.2, 0.3)^T$, $(0.1, 0.1, 0.1)^T$, $(0.1, 0.1,0.1)^T$, but EKF and DKF are $(y_1+ y_2+ y_3)+(S(t)+2)(x_2^2+x_3^2+50)$. DKF is $(T(t)-z_1)(x_2^2+x_3^2+50)-2$. The retrieved signal and error versus t are shown in Fig.3(c,d).

The results indicate that if Decrypter offers same KICs as encrypter and right DKF, then can correctly retrieve secret signal. It will be very hard for intruders to extract hidden information without knowing detail of DKF (or EKF) and KICs.

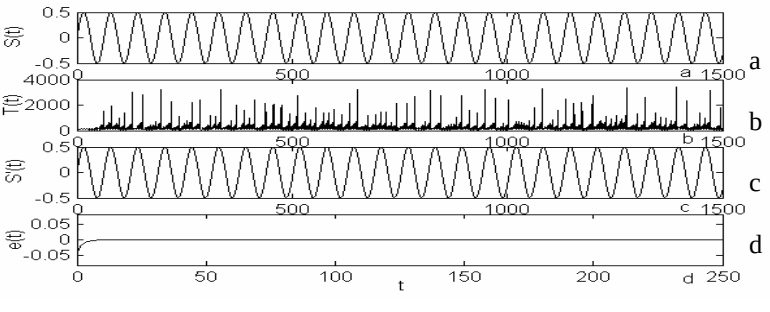


Fig. 2. Results of with right EKF and KICs

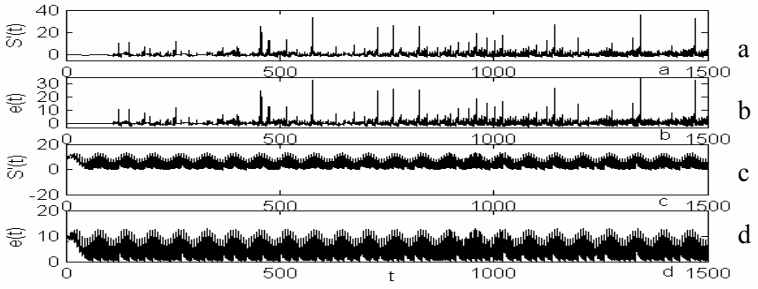


Fig. 3. Results of with incorrect EKF or KICs

4 Conclusion

With choosing appropriate EKFs, secret signal is "syncretised" into compound chaotic signal, so transmitted signal is complex. Numerical simulations show that our secure communication system is very sensitive to DKFs and KICs mismatch.

In next work, we will give a strong cryptanalysis of those methods and test robustness of the system to additive channel noise.

Acknowledgements

This work is supported by Science and Technology Foundation of Chongqing Education Commission. (Project No: KJ050802) and Chongqing Normal University Doctor Science and Technology Foundation (05XLB006).

References

1. L. K. Pecora and J. L. Carroll, "Synchronization in chaotic systems,"*Phy. Rev. Lett.*, vol. 64, pp. 821–824, 1990.
2. G. Heidari-Bsteni and C. MCGillem, "A chaotic direct-sequence spread spectrum communication system," *IEEE Trans. Commun.*, vol. 42, pp.1524–1527, Feb./Mar./Apr. 1994.
3. O. Gonzales, G. Han, J. Gyvez, and E. Sanchez-Sinencio, "Lorenz-based chaotic cryptosystem: a monolithic implementation," *IEEE Trans Circuits Syst. I*, vol. 47, pp. 1243–1247, Aug. 2000.
4. Naoki Masud, Kazuyuki Aihara. Cryptosystems With Discretized Chaotic Maps. IEEE Trans Circuit and Systems,2002,49(1) : 28-40
5. Huawei Ruan , Tongyan Zhai, Edwin Engin Yaz. Achaotic secure communication scheme with extended Kalman filter based parameter estimation. IEEE, 2003, 404-408.
6. Sh. B., Sq. W., Hq. Y, Synchronization of a unified chaotic system and the application in secure communication. Physica D 164 (2002) 45–52.

Studies on Neighbourhood Graphs for Communication in Multi Agent Systems

Andreas Goebels

International Graduate School of Dynamic Intelligent Systems
Knowledge Based Systems, University of Paderborn, Germany
swarmgroup@upb.de

Abstract. This paper addresses a special type of graph, the k-neighbourhood graph, for the usage in huge multi agent systems. It can be used to establish slim communication structures in extensive groups of agents as they are present e.g. in swarm applications. We will prove some properties of k-neighbourhood graphs in two- and three-dimensional Euclidean space, i.e. we show that the maximum number of incoming connections per agent is limited to a value independent from the overall number of agents in the system. For the two-dimensional case we prove a maximum in-degree of $6 \cdot k$. Furthermore, for agents interacting in three dimensions an upper and a lower bound for this value is presented.

1 Introduction

In several multi agent or swarm intelligence systems[1][2] it is shown that inter-agent communication can improve the performance of the whole system(see e.g. [3]). Several approaches exist that apply different techniques and concepts to establish such communication structures. These techniques can be divided into two main areas.

The first one contains all methods that offer global communication for each agent in the group. This can be realized by centralized techniques, e.g. the usage of a black- or whiteboard[4], or by allowing each agent to communicate directly with any other in some kind of complete communication structure[1]. When dealing with large numbers of agents as they are present in swarms or swarm like systems, the centralized approaches have some disadvantages. The blackboard can become a bottleneck and is a single point of failure. The complete communication structure shows some problems in such large settings, too, since the amount of messages grow very fast in such systems.

The second area is more applicable to large groups of agents since it allows only communication in a local area, mostly defined by a maximum communication radius[5] or by limiting communication to a special direction, e.g. defined by a cone. Such latter graphs are named after the author Yao-graphs[6]. These systems need a fixed, environment- or problem-dependent radius. A radius which is too big will produce similar problems as the global communication concepts entail because there can be too many communication partners. A radius which is too small can lead to a system with no communication (and hence no information exchange) at all. Other concepts in this area are e.g. *c-spanners*. These are graphs satisfying that between any two vertices there exists a path of length at most c in geometric distance[7].

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 456–465, 2006.

Our approach is related to the local communication concepts. The difference to the ideas we mentioned before is that we allow each agent to adjust its communication radius individually. Compared to Yao graphs, we do not claim that our neighbours are in different sectors. In detail, each agent will increase its radius until a fixed number k of other agents are accessible, wherever they may be located. We model our system by transforming the agents and communication lines into a more abstract geometrical graph structure. In such a graph, each agent is represented by a node and each communication line is an edge. In the remaining text we will use the terms *node* and *agent* in parallel. The arising graph is denoted as a k-neighbourhood graph (kNG). We used such graphs successfully in some coordination problems for agents (e.g. in [8]). In [9], we give an example for learning appropriate sizes of the k-neighbourhood for each agent individually when handling a multi objective optimization problem.

The kNG has been examined in [10], but there Eppstein considered only agents interaction in the two-dimensional space having an out-degree of $k = 1$.

The remaining part of this paper is structured as follows. In the next section, we give a formal definition of the k-neighbourhood graph and introduce some notations. The two-dimensional in-degree proof is layd out in section 3 and the bound for three dimensions in section 4. In the end, we conclude this paper and give ideas for future research and possible applications.

2 Notation

In this section, we formally define the k-neighbourhood graph and several functions or terms we use in this paper. First of all, we consider a set of n agents denoted by $\mathcal{A}$ with $n = |\mathcal{A}|$. We assume that no two agents are located on the same position in space (distance $\delta(a, a') > 0 \ \ \forall a, a' \in \mathcal{A},\ a \neq a'$).

Definition 1. *The function $\delta(p_1, p_2)$ defines the distance between two points in a three dimensional Euclidean space. This function works in the same way if we consider two arbitrary agents a and a', then $\delta(a, a')$ returns the distance between the positions of two agents*

$$\delta : \mathcal{A}\ x\ \mathcal{A} \rightarrow \mathbb{R} : \delta(a, a') = |a, a'| = \sqrt{(a_x - a'_x)^2 + (a_y - a'_y)^2 + (a_z - a'_z)^2}$$

Definition 2. *The* k-neighbourhood $\aleph_a$ *of an agent $a \in \mathcal{A}$ is defined to contain the k nearest agents.*

$$\aleph_a = \{a_1, ..., a_k\} \textit{ such that } \forall a' \in (\mathcal{A} \setminus \aleph_a) :\ \delta(a', a) \geq \delta(a_i, a)(i = 1, ..., k)$$

Definition 3. *We consider the graph $G = (V, E)$ created when connecting each of the n nodes ($n = |A|$) with its k nearest nodes. These are defined by the k-neighbourhood.*

$$V = A; E = \{(a, b) | a \in \mathcal{A}, b \in \aleph_a\}$$

We call this graph G a k-neighbourhood graph (kNG). If the edges are directed, the graph is a directed-kNG.

In [11], there is an algorithm that can compute such graphs in $O(k \cdot n \log n)$ and in [12] the algorithm needs $O(k \cdot n + n \log n)$.

3 Graph Properties 2D

In this section, we will provide some theoretical insight into the introduced k-neighbourhood graphs. It will be proven with geometrically properties that the number of incoming connections to an arbitrary node is independent from the overall number of agents and bounded by $6 \cdot k$. We start with a proof for $k = 1$ and extend this one for arbitrary values of k. In [10] there is another idea for this case, but we need this approach to show the more general relation for arbitrary k.

Theorem 1. *In a two-dimensional space the maximum number of incoming connections to an arbitrary agent $a_M \in A$ in a 1-neighborhood graph is exactly 6.*

We begin with a lemma proving that it is possible to locate 7 agents in such a way that one agent has 6 incoming connections.

Lemma 1. *For special values of n, agents can be located in such a way that one agent has 6 incoming connections if the agents are connected by a 1-neigbourhood graph.*

Proof. We arrange $(n-1)$ agents on the edges of a regular polygon with the remaining agent a_M in the centre of the polygon. We denote the distance between each agent and the agent a_M with r and the distance between two adjacent agents with s. We have the law of cosine

$$s^2 = 2r^2 - 2r^2 cos\left(\frac{2\pi}{n-1}\right) \Leftrightarrow r = \pm \frac{1}{\sqrt{2 - 2cos\left(\frac{2\pi}{n-1}\right)}} \cdot s \Leftrightarrow r = \Delta \cdot s$$

(We can ignore the negative value for the radius r)
For Δ smaller than 1, r is smaller than s. Hence, there will be a connection from an agent in the polygon with the agent a_M. This inequality is only true for $(n-1) < 6$ or $n < 7$. In the case $n = 7$ the distance from one agent to the agent a_M in the middle of the polygon is the same as the distance to its two adjacent polygon neighbours, therefore the agents can still choose the agent in the middle and the in-degree of a_M can be 6. □

Now we start with the proof for theorem 1. Therefore, we divide the space around agent a_M in s sectors (or cones) S_i with equal size. We choose for each sector the agent n_i with the lowest distance to agent a_M. An agent a will be in the sector S_i if it is situated directly inside a sector (see figure 1(a)). If an agent is located on the border between two sectors, it will belong only to the right one. Now we consider an arbitrary sector S_i with the agent n_i that lies nearest to a_M and we show that there can be at most one connection to a_M. Let A_i be the set of agents that are in this sector S_i with $n_i, a_M \notin A_i$. For all agents $a \in A_i$ it holds that $\delta(a, a_M) > \delta(a, n_i)$ if we have a sufficient large number s of sectors. This is proven in the following Lemma.

Lemma 2. *Let S be a sector originating at a_M with an angle α in a two-dimensional space and let n_i and a be two points in S with $|a_M, n_i| \leq |a_M, a|$. Then $|a, n_i| < |a, a_M|$ for $\alpha \in [0, \frac{\pi}{3}[$.*

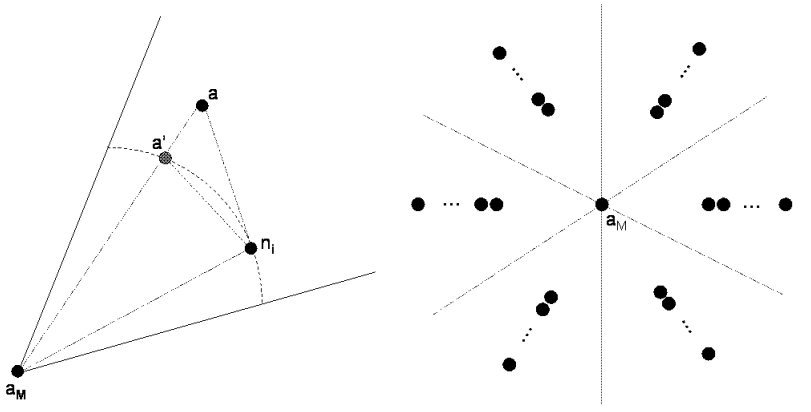


Fig. 1. A sketch of one sector that contains n_i and a and the order of all nearest agents n_{x_y} in a way that they can choose a_M as one of the k nearest agents

Proof. First we define an imaginary point a' that lies on the line from a_M to a with the same distance to a_M as n_i has (see figure 1(a)). The law of cosine states:

$$|a', n_i|^2 = |a_M, n_i|^2 + |a_M, a'|^2 - 2 \cdot |a_M, n_i| \cdot |a_M, a'| \cdot cos(\beta)$$

Because of the construction of point a', it holds that

$$|a', n_i|^2 = 2 \cdot |a_M, n_i|^2 - 2 \cdot |a_M, n_i|^2 \cdot cos(\beta)$$

We use a trigonometric transformation and obtain

$$|a', n_i| = 2 \cdot sin\left(\frac{\beta}{2}\right) \cdot |a_M, n_i| \tag{1}$$

Applying the triangle inequality we know that

$$|a, n_i| \leq |a', n_i| + |a, a'| \tag{2}$$

and, of course, it holds that

$$|a, a'| = |a_M, a| - |a_M, a'| = |a_M, a| - |a_M, n_i| \tag{3}$$

So, if we combine (1), (2) and (3):

$$|a, n_i| \leq |a_M, a| - \underbrace{\left(1 - 2 \cdot sin\left(\frac{\beta}{2}\right)\right) \cdot |a_M, n_i|}_{=\Delta}$$

For $\beta \in [0, \frac{\pi}{3}[$ Δ is always > 0, therefore $|a, n_i| < |a_M, a|$. □

Lemma 3. *There can be at most 6 sectors with one agent that chooses a_M as communication partner.*

Proof. Assume we have $s > 6$ sectors with an agent n_i[1] each. There is at least one pair of agents (n_{i_1}, n_{i_2}) with an angle γ between the lines from a_M to each of it with $\gamma < \frac{2\pi}{6}$. In lemma 2 we have proven that for such a setting only one of these agents can choose a_M as a communication partner. This is a contradiction to our assumption, hence s has to be ≤ 6. □

Now we have all the information we need to prove theorem 1:

Proof. If we choose a number of sectors that is equal to 6, we can define β to be an angle $< \frac{\pi}{3}$ because of the definition that points on the border between two sectors belong only to the rightmost one and $\beta \leq \alpha < \frac{\pi}{3}$ (due to construction). Thus we know that for all agents $a \in A_i$ the agent n_i is located nearer to a than the agent a_M (see Lemma 2). The in-degree can no more be increased, it's maximum value is 6. □

In the next theorem we will expand this result for arbitrary k's, the proof is very similar to the one of theorem 1.

Theorem 2. *In a two-dimensional space the maximum number of incoming connections to an arbitrary agent $a_M \in \mathcal{A}$ in a k-neighbourhood graph is $6 \cdot k$.*

Proof. Again we divide the space around agent a_M into 6 uniform sectors $S_1, ..., S_6$. Without loss of generality we consider one of these sectors denoted by S_i. Now we consider the k nearest agents $n_{i_1}, ..., n_{i_k}$ with

$$\delta(a_M, n_{i_1}) \leq \delta(a_M, n_{i_2}) \leq ... \leq \delta(a_M, n_{i_k})$$

in each sector to the agent a_M (see figure 1(b)).

Let A_i contain the remaining agents in this sector $A_i = \{a_j | a_j \text{ located in } S_i, a_j \neq n_{i_l} \forall l \in \{1, ..., k\}\}$. If each of these agents n_{i_x} can be located in the sector in such a way that the agent a_M is among the k nearest ones, the agent a_M has an in-degree of $6 \cdot k$. In figure 1(b) one can find an example how an in-degree of $6 \cdot k$ can be realized. For construction we use the same idea as we did in lemma 1.

Now we can argue as we did in theorem 1. If we consider any agent $a \in A_i$, the k nearest agents are among other agents in A_i (or maybe other agents in other sectors), or at least the k agents n_{i_k}. The reason is the same as in the proof for theorem 1. There is no chance to chose the agent a_M as the nearest one, the n_{i_x} agents are always closer. Therefore, the in-degree of the agent a_M could never become higher than $6 \cdot k$. □

4 Graph Properties 3D

In this section, we will concentrate on the three-dimensional space and will determine a lower and an upper bound for the in-degree. Therefore, we start with giving an example how an in-degree of $12 \cdot k$ can be constructed and will then prove that an in-degree greater than $60 \cdot k$ cannot occur.

[1] This is again the agent with the least distance to a_M in sector i.

Theorem 3. *In a three-dimensional space the in-degree for an arbitrary agent a_M can be equal to $12 \cdot k$.*

Proof. For this lower bound proof we construct an ikosaeder $\mathcal{I}$ (see figure 2(a)) with side length l, vertex set E = $\{e_1 ... e_{12}\}$ and a_M as the centre. The outer sphere perimeter radius of $\mathcal{I}$ is

$$R = 0.25\sqrt{10 + 2\sqrt{5}} \cdot l \approx 0.95106 \cdot l < l$$

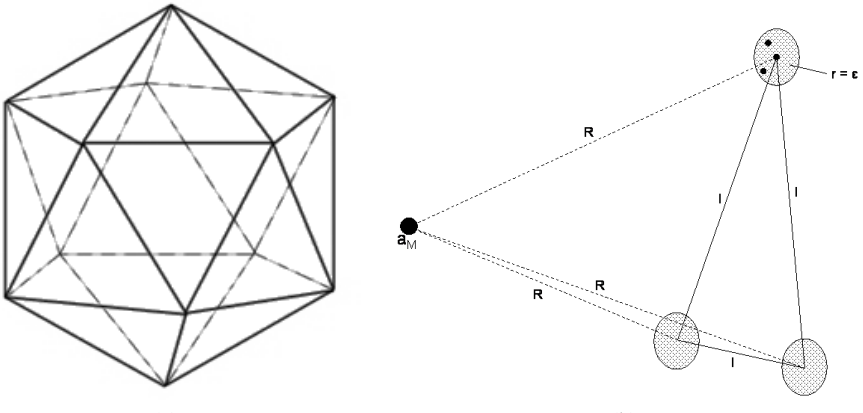
We distribute $12 \cdot k$ agents in such a way that on each vertex there are exactly k agents in a small ϵ environment located. Now we show for any k that such an agent positioning results in an in-degree of exactly $12 \cdot k$ for agent a_M.

For $k = 1$, the analysis is very simple. If we locate agents on all vertexes of $\mathcal{I}$ (i.e. $\epsilon = 0$), the agent a_M in the middle will have an in-degree of 12 because for each agent, a_M is the nearest adjacent agent. This is due to the geometrical properties of an ikosaeder. Each vertex has a maximum distance of R to the centre point, maybe less. At the same time, the side length, i.e. the distance between adjacent vertexes, is l and this is greater than R. Hence, the agent a_M in the centre is closer than any agent on the ikosaeder vertexes. Therefore, a_M is chosen by all 12 agents and has an in-degree of $12 \cdot 1$.

This idea can be enhanced for arbitrary k. If we consider $k > 1$, we can locate k agents in an $\epsilon \in \mathbb{R}_0^+$ environment around each vertex e_x. This environment can be seen as a sphere with radius ϵ and e_x in the center. We denote the agents around an arbitrary vertex e_i by $A_i = \{a_1, ..., a_k\}$. By definition, the distance from vertex e_i to any $a \in A_i$ is bounded by ϵ (i.e. $\delta(e_i, a) \leq \epsilon$).

If we now choose

$$\epsilon < \frac{1}{3} \cdot (l - \frac{1}{4}\sqrt{10 + 2\sqrt{5}} \cdot l) \Leftrightarrow R < l - 3 \cdot \epsilon$$



(a) (b)

Fig. 2. Figure 2(a) visualizes the geometrical shape of an ikosaeder. It consists of 20 triangles of the same size and 12 vertexes. In figure 2(b), one of these regular triangles and the ϵ-environment, where the agents are located, is schematicly displayed.

and consider the ikosaeder with side length l, it holds for each agent $a \in A_i$ that $(k-1)$ agents are in the same ϵ-neighbourhood (the agent will not choose itself as a neighbour). Due to the location of the ϵ-neighbourhood sphere we know that $|a, a_M| \leq (R+\epsilon) \ \forall a \in A_i$.

At the same time, the distance between any agent $a \in A_i$ and an arbitrary agent $b \in A_j$ in the ϵ-environment of edge $e_j \ \ (i \neq j)$ is: $|a, b| \geq (l - 2 \cdot \epsilon)$. Therefore, the following inequality holds:

$$|a, a_M| \leq (R+\epsilon) < (l - 3 \cdot \epsilon) + \epsilon = (l - 2 \cdot \epsilon) \leq |a, b|$$

Hence, each agent in an arbitrary ϵ-environment will choose the remaining (k-1) agents and then the agent a_M as communication partners, thus a_M has in-coming connections from all $12 \cdot k$ agents. □

Definition 4 (3-face Ikosaeder (3F-Ikosaeder)). *We introduce a new geometrical figure that is similar to the ikosaeder (see figure 2(a)). In the 3F-Ikosaeder, we divide each of the 20 faces of the ikosaeder with side-length l into three faces of equal size. Therefore, we consider a single face with the vertexes A, B and C. These three vertexes form an equilateral triangle (with the side length l). The median lines from each side cross in a unique centroid point S (see figure 3(b)). Now we can construct three faces with the same area as shown in figure 3(a). This quadrangle has the side length $\frac{1}{6} \cdot \sqrt{3} \cdot l$ and $\frac{l}{2}$, the diagonals have the length $\frac{1}{\sqrt{3}} \cdot l$ and $\frac{l}{2}$.*

The resulting geometric object has obviously $20 \cdot 3 = 60$ faces. An example is shown in figure 4(a). From the geometrical properties of the Ikosaeder we know the in-sphere radius

$$r = (1/12) \cdot \sqrt{3} \cdot (3 + \sqrt{5}) \cdot l \approx 0.755761 \cdot l$$

and the same circum-sphere radius $R = 0.25\sqrt{10 + 2\sqrt{5}} \cdot l$ as the Ikosaeder presented before.

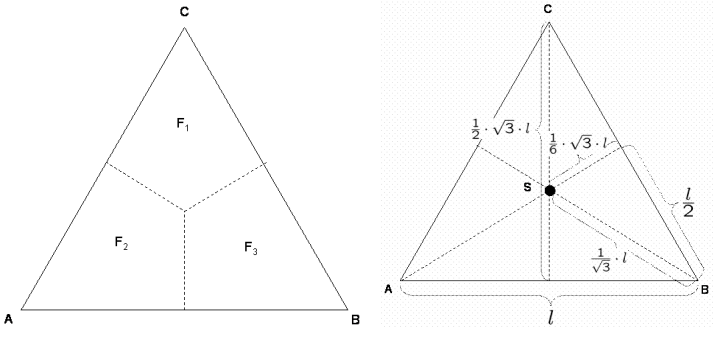
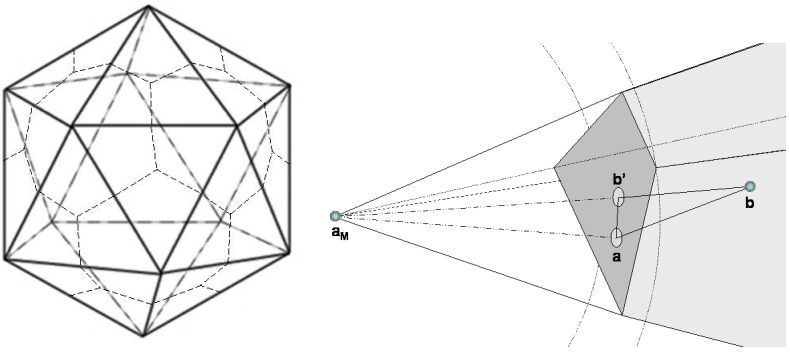


Fig. 3. Figure 3(a) shows the construction of the single faces of the 3F-Ikosaeder. In figure 3(b), the geometrical properties of these faces are presented. The side length of the whole 3F-Ikosaeder is l.



(a) (b)

Fig. 4. The 3F-Ikosaeder and a schematic figure of the proof for theorem 4

Theorem 4. *In a three-dimensional space the maximum number of incoming connections to an arbitrary agent $a_M \in \mathcal{A}$ in a* 1*-neighbourhood graph can not be greater than* $60 \cdot 1$. *This is an upper bound.*

Proof. We start with the construction of a 3F-Ikosaeder $\mathcal{I}$ as defined before. The side length of $\mathcal{I}$ is variable and denoted by l. The centre of $\mathcal{I}$ is the agent a_M. The whole space around a_M can now be divided into 60 cone-like structures $\mathcal{S} = \{S_1, ..., S_{60}\}$. Each S_i can be seen as the extension of a pyramid, with a_M as apex and one face of $\mathcal{I}$ as basis. Without loss of generality, we consider an arbitrary segment $S \in \mathcal{S}$.

In S, we consider the agent $a \in \mathcal{A}$ with the shortest distance to a_M. We modify now the side length of $\mathcal{I}$ in such a way that a lies directly in a face of $\mathcal{I}$. An example is shown in figure 4(b).

Now we will prove that the distance from an arbitrary agent b that lies in S to this agent a is always smaller than the distance from b to agent a_M. Hence, the in-degree of a_M is bounded by the number of faces of the ikosaeder.

Therefore, we have to show: $|b, a| < |b, a_M|$

We know that all points on the surface of an 3F-Ikosaeder face lie between or on the in- and the circum-sphere, therefore we know that

$$r \leq |a_M, a| \leq R \tag{4}$$

We can add the projection of b[2] (denoted by b') onto the face into this relation, then we can restate our goal to $|b, a_M| = |a_M, b'| + |b', b| > |b, a|$ If we add here equation (4), we can consider the minimum value for $|a_M, b'|$ and replace this part by its minimum value r, hence it is enough to show that

$$r + |b', b| > |b, a| \tag{5}$$

Because the triangle inequality holds, we know that

$$|b, a| \leq |b', b| + |b', a| \Leftrightarrow |b', b| \geq |b, a| - |b', a| \tag{6}$$

[2] The projection of b onto an ikosaeder face is the intersection point of the connection between b and a_M and the face itself.

If we combine (5) and (6), we get

$$r + |b, a| - |b', a| > |b, a| \Leftrightarrow r > |b', a|$$

We know from our definition of the 3F-Ikosaeder that the longest distance d in one face is

$$d = \frac{1}{\sqrt{3}} \cdot l \approx 0.57735 \cdot l$$

Therefore, the distance from b' to a is bounded by d and we can say that

$$r = \frac{1}{12} \cdot \sqrt{3} \cdot (3 + \sqrt{5}) \cdot l > \frac{1}{\sqrt{3}} \cdot l = d$$

This statement is true for arbitrary $l > 0$, thus we have shown that in each sector only the agent that is closest to a_M can choose this agent as communication partner. In each cone the agent a is some kind of barrier for further connections of other agents with a_M. Hence, we have proven an upper bound of $60 \cdot 1$ for the number of incoming connections.

Now, we will prove the upper bound for incoming connections for arbitrary k.

Theorem 5. *In a three-dimensional space the maximum number of incoming connections to an arbitrary agent $a_M \in \mathcal{A}$ in a k-neighbourhood graph cannot be greater than $60 \cdot k$.*

Proof. Assume, we have more than $60 \cdot k$ agents with connections to the agent a_M. If we again construct an 3F-Ikosaeder around a_M, then there has to be at least one cone-like structure with a_M as apex (cf. figure 4(b)) with $k' > k$ agents that have connections to a_M. Now we consider the $(k+1)$-nearest agent to a_M. With the same argumentation as we did in the proof for theorem 4 we can show that this agent will choose as its communication partners each of the k-nearest agents to a_M prior. Hence, there is no possibility left to choose a_M as communication partner and this is a contradiction to our assumption. □

If we now combine the results from Theorem 3 and Theorem 5, we have a lower and an upper bound for the maximum number of communication partners for arbitrary agents in a kNG. This number $m \in \{12 \cdot k, ..., 60 \cdot k\}$ is independent from the overall number of agents.

5 Conclusion and Future Work

In this paper we were engaged with a special kind of graph. We have proven that in such graphs with k outgoing edges to the nearest agents the in-degree of each agent is independent from the overall number of agents. The maximum value could be determined exactly for two-dimensional interactions by $6 \cdot k$. For three-dimensional interactions, we could prove lower and upper bounds of $12 \cdot k$ and $60 \cdot k$ by the construction of a special geometrical structure called 3F-Ikosaeder.

In some experiments we compared the kNG communication structure with the classical and often used radial communication structure concerning energy costs. In experimental settings we found out that we can save a lot of energy (more than 40%) with only small values for k. At the same time, we do not need a parameter for such a communication radius. Our system can generate a connected graph with very small values for k. These experiments and results will be presented in an upcoming paper.

References

1. Ferber, J.: Multi-agent systems: An introduction to distributed artificial intelligence. Addison-Wesley, Harlow, England [u.a.] (1999)
2. Bonabeau, E., Dorigo, M., Theraulaz, G.: Swarm Intelligence - From natural to artificial Systems. Oxford University Press (1999)
3. Matarić, M.J.: Using communication to reduce locality in distributed multi-agent learning. In: Journal of Experimental and Theoretical Artificial Intelligence, special issue on Learning in DAI Systems, Gerhard Weiss, ed., 10(3) (1998) 357–369
4. Nii, H.P.: Blackboard systems, part one: The blackboard model of problem solving and the evolution of blackboard architectures. AI Magazine **7** (1986) 38–53
5. Gao, J., Guibas, L., Hershberger, J., Zhang, L., Zhu, A.: Geometric spanner for routing in mobile networks. (2001)
6. Eppstein, D.: Spanning trees and spanners. Technical Report Technical Report 96-16, University of California, Dept. Information and Computer Science (1996)
7. Schindelhauer, C., Volbert, K., Ziegler, M.: Spanners, weak spanners, and power spanners for wireless networks. In: Proceedings of ISAAC 2004, LNCS 3341, Springer (2004) 805–821
8. Goebels, A., Büning, H.K., Priesterjahn, S., Weimer, A.: Towards online partitioning of agent sets based on local information. In: Proceedings of the International Conference on Parallel and Distributed Computing and Networks (PDCN), Innsbruck, Austria (2005)
9. Goebels, A.: Learning useful communication structures for groups of agents. In: Proceedings of the IFIP Conference on Biologically Inspired Cooperative Computing, BICC'06, Santiago de Chile, Chile (2006)
10. Eppstein, D., Paterson, M.S., Yao, F.F.: On nearest-neighbor graphs. GEOMETRY: Discrete & Computational Geometry **17** (1997)
11. Vaidya, P.: An $O(n \log n)$ algorithm for the all-nearest-neighbors problem. In: Discrete and Computational Geometry (4). (1989) 101–115
12. Callahan, P.: Optimal parallel all-nearest-neighbours using the well-separated pair decomposition. In: Proceedings of the 34th IEEE Symposium Foundations of Computer Science. (1993) 332–340

Evolutionary Dynamics of an Asymmetric Game Between a Supplier and a Retailer*

Zhou Min and Deng Fei-qi

Institute of Systems Engineering, South China University of Tech., Guangzhou 510640, China

Abstract. There are restrictions of complete rationality and information symmetry, which is difficult to work in reality, in the traditional game theory. However, the evolutionary game theory, based on bounded rationality, can compensate these defects. In this paper, we set up an asymmetric model of the evolutionary game between a supplier and a retailer with asymmetric information. The evolutionary stable strategies and their premises were obtained with the replicator dynamics mechanism. The analysis in this paper has significance of explanation and direction for the supply chain management.

1 Introduction

Since the game theory generated in the 20th century, it rapidly became one of the major methods to the problems of supply chains. Cachon and Zipkin(1999)[1] proposed a model of an inventory game in a two-stage supply chain. Then Cachon(2001)[2] expanded it, he set up a model that a supplier and several retailers choose the re-ordering point. Gurdal(2002)[3] articulated a game model of a two-stage supply chain in which the retailing price is a linear function of the wholesaling price. LI Ling and DENG Fei-qi(2006)[4] gave detailed introduction of the cooperative game analysis on enterprises coalition.

Above mentioned game theory models are based on the traditional game theory, its prerequisite is complete rationality which calls for the perfect master of the game. But in reality, a supplier and a retailer usually cannot know each other's payoff function, and their rationality also will encounter various restrictions. As a consequence, attention has shifted to evolutionary game theory which is based on bounded rationality and emphasizes the dynamic evolving process.

This paper provides an asymmetric model of evolutionary game between a supplier and a retailer with asymmetric information, and get the evolutionary stable strategies (ESSs) of it with the replicator dynamics mechanism. To the best of our knowledge, this is the first time that evolutionary game theory is used to analyze the asymmetric model of upper and lower corporations in supply chains. Previous applications of evolutionary game theory are all based on the fixed parameters, the difference of this paper is to analyze the different ESSs in case of variable parameters.

* This work is supported by the NSFC under the project 60374023.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 466–469, 2006.

2 ESS and Learning Mechanism

In the bounded rational analytic frame of evolutionary game theory, a player's behavior is not optimal at first, instead, his decision is optimized by dynamic processes such as individuals' mutual imitation, learning and mutation. ESS is one of the core concepts in evolutionary game theory. The definition of an ESS requires that for all mutant populations R there exists an invasion barrier ε such that the original population P does better against the mixed population $(1-\eta)P+\eta P$ than R does for all $\eta \leq \varepsilon$[5]. That is to say a ESS is a strategy with which a whole population cannot be invaded by small set of mutant genotypes.

Another key in evolutionary game theory is to confirm the mode for players' strategy adjustment, which we call learning and evolutionary mechanism. The most concerned one is replicator dynamics, the thought of which is players within a population change strategy from one to another with rates being functions of the average payoff(fitness) of the strategies. We adopt this mechanism because it's suitable for the large group and random matching environment in the global supply chains which are the trend of modern supply chain management.

3 The Evolutionary Game Model Between a Supplier and a Retailer

Consider the principal-agent relationship between a supplier and a retailer. The supplier provides products to the retailer for sell. The retailer can choose the strategy: promotion or non-promotion, and doesn't know whether the supplier will inspect his action. Owing to asymmetric information, the supplier can't examine whether the retailer works hard or not, but can only find out the final sales amount, thus to choose his strategy: inspection or non-inspection. The decision-making processes have the features of asymmetry and bounded rationality, and decisions of the supplier and the retailer interact, so it's suitable to analyze the processes with evolutionary game theory.

Notes:

e_1,e_2 respectively stand for profits of the supplier and retailer from selling one unit of product,e_1,$e_2 > 0$; B stands for the fixed sales quantity if the retailer doesn't promote sales of the product, $B > 0$; c_s, c_r respectively stand for inspection cost of the supplier and sales promotion cost of the retailer, $c_s, c_r > 0$; c_p represents the punishment suffered by the retailer when he doesn't promote sales of the product but the supplier inspects, $c_p > 0$; α represents the sales quantity factor when the retailer promotes sales of the product but the supplier doesn't inspect, $\alpha > 1$; β represents the sales quantity factor when the retailer promotes sales of the product and the supplier inspects, $\beta > \alpha$.

Evolutionary game is played among suppliers and retailers groups, every time the game is played by one member of the former group and one member of the later one. Every side make use of replicator dynamics to adjust his strategy.

Following is the payoff matrix of the supplier and the retailer:

Supplier \ Retailer	Promotion	Non-Promotion
Inspection	$(e_1\beta B - c_s, e_2\beta B - c_r)$	$(e_1 B - c_s, e_2 B - c_p)$
Non-Inspection	$(e_1\alpha B, e_2\alpha B - c_r)$	$(e_1 B, e_2 B)$

Suppose x as the proportion of the suppliers using the "inspection" strategy in suppliers group, (1-x) as "non-inspection"; y as the proportion of the retailers using the "promotion" strategy in retailers group, (1-y) as "non-promotion".

Then the expected payoff of a supplier when choosing "inspection" is:

$$E_i = (e_1\beta B - c_s)y + (e_1 B - c_s)(1-y) = e_1 B(\beta - 1)y + e_1 B - c_s \tag{1}$$

It follows that the expected payoff of suppliers group is:

$$\overline{E_s} = xE_i + (1-x)E_{\bar{i}} = e_1 B(\beta - \alpha)xy - c_s x + e_1 B(\alpha - 1)y + e_1 B \tag{2}$$

Thus we get the replicator dynamics equation of suppliers group:

$$\dot{x} = x(E_i - \overline{E_s}) = x(1-x)[e_1 B(\beta - \alpha)y - c_s] \tag{3}$$

Similarly the replicator dynamics equation of retailers group is:

$$\dot{y} = y(1-y)\{[e_2 B(\beta - \alpha) + c_p]x + e_2 B(\alpha - 1) - c_r\} \tag{4}$$

In above nonlinear system, order $\dot{x} = 0$, $\dot{y} = 0$, we get five equilibriums: (0,0), (0,1), (1,0), (1,1), $(\frac{c_r - e_2 B(\alpha-1)}{e_2 B(\beta-\alpha)+c_p}, \frac{c_s}{e_1 B(\beta-\alpha)})$.

As can be seen in reference[6], when a equilibrium of the replicator dynamics equations is an evolutionary equilibrium (equals to the locally asymptotically stable point in dynamic systems), it is an ESS.

So we can judge whether the five equilibriums are ESSs via analyzing Jacobian Matrix of this system. We get the Jacobian Matrix:

$$J = \begin{bmatrix} [e_1 B(\beta - \alpha)y - c_s](1-2x) & x(1-x)[e_1 B(\beta - \alpha)] \\ [e_2 B(\beta - \alpha) + c_p]y(1-y) & e_2 B(\beta - \alpha)x + c_p x + e_2 B(\alpha - 1) - c_r](1-2y) \end{bmatrix} \tag{5}$$

The equilibriums fitting $det(J) > 0$ and $tr(J) < 0$ are asymptotically stable, thus they are ESSs of the game[7]. Thus we conclude that:

i When $e_2 B(\alpha - 1) - c_r < 0$, there is one ESS (0,0), the strategy profile which the supplier and retailer adopted will converge to (Non-inspection, Non-promotion);

ii When $e_1 B(\beta - \alpha) - c_s < 0$ and $c_r - e_2 B(\alpha - 1)$, there is one ESS (0,1), the strategy profile will converge to (Non-inspection, Promotion);

iii When $c_s - e_1 B(\beta - \alpha) < 0$ and $c_r - c_p - e_2 B(\beta - 1) < 0$, there is one ESS (1,1), the strategy profile will converge to (Inspection, Promotion);

iv When $e_2 B(\alpha - 1) - c_r < 0$, $c_s - e_1 B(\beta - \alpha) < 0$ and $c_r - c_p - e_2 B(\beta - 1) < 0$, there are two ESS (0,0)and (1,1), the strategy profile will converge to (Non-inspection, Non-promotion) or (Inspection, Promotion) according to different initial strategy profiles.

4 Conclusion

Game theory has become one of the main methods to solve the problems in supply chains. But the former articles are generally based on complete rationality and strictly require the players to master correct knowledge and abilities of analysis and judgement for the structure of the game, and the condition of every player. Actually in the real market economy, even if the upper and lower streams are in cooperative partner relationship, it's still difficult to guarantee absolute information symmetry. So we generate an asymmetric information evolutionary game model between a supplier and a retailer, and analyze the ESSs of the game with the replicator dynamics learning mechanism. The conclusion is that the ESSs in this evolutionary game model are influenced by the parameter values and the initial strategy profile.

This paper can clearly and effectively explain the strategy evolutionary processes of upper and lower enterprises in supply chains. It also gives theoretic basis for the national and government to interfere supply chain operations.

References

1. Cachon G, Zipkin P: *Competitive and Cooperative Inventory Policies in a Two-stage Supply Chain*. Management Science. 45 (1999) 936-953
2. Cachon G. Stock: Wars: *Inventory Competition in a Two Echelon Supply Chain*. Operation Research. 49 (2001) 658-674
3. Gurdal E, Griffin P M: *Supplier and Buyer Driven Channels in aTwo-stage Supply Chain*. IIE Transactions. 34 (2002) 691-700
4. LI Ling, DENG Fei-qi. *Research on Enterprises Coalitional Games*. The First International Conference on Complex Systems and Applications, Huhhot, Inner Mongolia, China (2006). (in press)
5. Jorg Oechssler, Frank Riedel: *On the Dynamic Foundation of Evolutionary Stability in Continuous Models*. Journal of Economic Theory. 107 (2002) 223-252
6. Ross Cressman: *The Stability Concept of Evolutionary Game Theory*. Berlin Heidelberg: Springer-Verlag (1992)
7. Ronald Shone: *Economic Dynamics*. China Renmin University Press (2003)
8. Taylor. P. D, L. B. Jonker: *Evolutionary Stable Strategy and Game Dynamics*. Math Biosci. 40 (1978) 145-156
9. Friedman D: *Evolutionary Games in Economics*. Econometrica. 59 (1991) 637-666

A Genetic Algorithm-Based Double-Objective Multi-constraint Optimal Cross-Region Cross-Sector Public Investment Model*

Tian Lei[1], Liu Lieli[1], Han Liyan[1], and Huang Hai[2]

[1] School of Economics and Management
Beihang University, Beijing 100083, P.R. China
tianlei@sem.buaa.edu.cn
[2] School of Computer Science and Engineering
Beihang University, Beijing 100083, P.R. China
huanghai@vrlab.buaa.edu.cn

Abstract. An optimal public investment model with two objective functions considering efficiency & equity and several constraints such as taxes and capital transfer loss are established by dividing public & private sectors and relaxing several original hypotheses respectively. And the objective functions and constraints are handled to adapt the model into the double-objective multi-constraint programming model suitable for genetic algorithm-based solution. Then encoding and decoding approaches are designed. Finally a case study is carried out to validate the proposed model and the GA-based solution.

1 Introduction

Optimal public investment model is evolved from optimal economic growth model. In nature, it is the optimal economic growth model focusing on optimal public investment. Research on optimal economic growth model emerged in 1920's and reached a peak in 1980's. In recent years, academics from many countries have studied optimal public investment from many specific aspects. For example, Gaube [1] researched public investment and income taxation, Basu, etc. [2] researched cross-country economic growth and inequality correlation, Brecher, etc. [3] researched dynamic stability in a two-country model of optimal economic growth and international trade, Van, etc. [4] researched recursive utility and optimal economic growth, Jouini [5] researched convergence of utility Functions and optimal strategies, Aurell [6] researched optimal growth strategies, Saglam [7] researched optimal economic growth model and the lagrange multiplier, and Gómez [8] researched optimal fiscal policy in Uzawa-Lucas model with externalities, etc. However, there are little literature on establishment of and solution to the whole optimal public investment model. In this paper, optimal public investment is studied from a spatial perspective, i.e. capital flows and labor flows are caused by different marginal capital productivity,

* This paper is supported by the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20050006025).

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 470–479, 2006.

marginal labor productivity, production function, income and other factors of different regions, which will further affect public investment of the next period. Therefore, it is of realistic significance to study optimal cross-region public investment.

2 Review of Optimal Economic Growth Model

Wu, Jih-Hwa [9] carried on extensive work on optimal economic growth model and its extensions to cross-region investment. Since Ramsey published his seminal work about optimal economic growth [10] in Economic Journal in 1928, much literature has come out. In neo-classical optimal economic growth model under aggregative close economy, the time path of consumption per capita should be selected to solve the following single-objective maximal problem:

$$\begin{array}{ll} Max & W = N\int_{T_0}^{T} e^{-\mu(t-T_0)}U(c(t))dt \\ Subject\ to & k'(t) = f(k(t)) - c(t) - \rho k(t) \\ & k(t_0) = k_0 \\ & 0 \le c(t) \le f(k(t)), \end{array} \tag{1}$$

among which W is social welfare; the individual utility function U at time t is the function of $c(t)$; N is the gross population; μ is exponential discounting factor; $c(t)$ is piecewise continuous; output $Y(t)$, consumption $C(t)$, and investment $I(t)$ are all functions of t; the current capital stock depreciates at a constant rate γ; suppose $K(t)$ is the capital stock at t, and then:

$$I(t) = K'(t) + \gamma K(t) \ . \tag{2}$$

Divide $K(t)$ by simple labor, i.e. $k(t) = K(t)/L(t)$; suppose labor supply grows at a constant exponential rate θ, $i(t) = k'(t) + (\gamma + \theta)k(t)$, let $\rho = \gamma + \theta$, so (2) can be rewritten as $k'(t) = f(k(t)) - c(t) - \rho k(t)$.

Solutions to (1) have been addressed in much economic growth literature.

3 Establishment of Optimal Regional Public Investment Model

Suppose there are n different regions. Part of the income Y_i of region $i(i = 1, ..., n)$ is saved for investment. And the central government distributes the investment fund to every region according to the principle of maximizing the total regional incomes. In this section, the cross-region cross-sector production function is established first, then the objective functions and constraints of optimal regional public investment model are discussed in turn, and the double-objective, multi-constraint, cross-region, and cross-sector optimal regional public investment model will be established finally.

3.1 Cross-Region Cross-Sector Cobb-Douglas Production Function

First, the assumption that aggregative production function is homogenous production function of degree one is relaxed as Cobb-Douglas production function is adopted. According to the simple model of Mankiw, Romer, and Weil [11], Cobb-Douglas production function is:

$$Y(t) = F(K(t), L(t)) = AK(t)^{\alpha} L(t)^{\beta} \ , \tag{3}$$

among which A represents technical factor. Suppose $0 < \alpha < 1$, $0 < beta < 1$, which indicate the marginal outputs of manufacturer's capital and simple labor are declining. Theoretically, $(\alpha + \beta)$ may be larger than 1, equal to 1, or smaller than 1, which figure increasing returns to scale, constant returns to scale, and decreasing returns to scale.

The regional production functions and economic structures are different, in the same region, the industrial sector may be in the stage of constant returns to scale, but the service sector may be in the stage of increasing returns to scale, so the aggregative production function should be break down by region and sector.

Suppose there totally are m sectors in every region. Then the Cobb-Douglas production function of sector j in region i is:

$$Y_{ij}(t) = F_{ij}(K_{ij}(t), L_{ij}(t)) = A_{ij} K_{ij}(t)^{\alpha_{ij}} L_{ij}(t)^{\beta_{ij}} \ . \tag{4}$$

Then the Cobb-Douglas production function of region i is:

$$Y_i(t) = \sum_{j=1}^{m} F_{ij}(K_{ij}(t), L_{ij}(t)) = \sum_{j=1}^{m} A_{ij} K_{ij}(t)^{\alpha_{ij}} L_{ij}(t)^{\beta_{ij}} \ . \tag{5}$$

And then the aggregative Cobb-Douglas production function is:

$$Y(t) = \sum_{i=1}^{n} F_i(K_i(t), L_i(t)) = \sum_{i=1}^{n} \sum_{j=1}^{m} A_{ij} K_{ij}(t)^{\alpha_{ij}} L_{ij}(t)^{\beta_{ij}} \ . \tag{6}$$

3.2 Cross-Region Cross-Sector Public Investment Objective Functions

The aims of cross-region cross-sector public investment may include but not limited to final total income maximization and cross-region equity maximization. The control variables include rate of savings, tax rates, public investment shares of regions and sectors, etc.

Cross-Region Cross-Sector Final Total Income Maximization. Friesz and Luque (1985) were the first to propose the income objective function considering relative importance of different regions. And in this section, the relative importance of different sectors will also be considered. Suppose the income of sector j in region i at the final time T is $Y_{ij}(T)$. Then the total income of region

i at the final time T is $\sum_{j=1}^{m} Y_{ij}(T)$. Considering the relative importance of different sectors, a weight can be added in front of the income of every sector, and then the weighted total income of region i at the final time T is $\sum_{j=1}^{m} \xi_{ij} Y_{ij}(T)$, among which ξ_{ij} is the weight of sector j in region i, and $\sum_{j=1}^{m} \xi ij = 1$. Therefore the cross-region total income at the final time T is $\sum_{i=1}^{n} \sum_{j=1}^{m} \xi_{ij} Y_{ij}(T)$. Similarly, a weight can also be added in front of the regional incomes before summing up, i.e.: $\sum_{i=1}^{n} \sum_{j=1}^{m} \omega_i \xi_{ij} Y_{ij}(T)$, among which ω_i is the weight of region i, and $\sum_{i=1}^{n} \omega_i = 1$. If the final total income is used to represent total welfare, then the maximization objective of cross-region final total income is:

$$Max\ J_1 = \sum_{i=1}^{n} \sum_{j=1}^{m} \omega_i \xi_{ij} Y_{ij}(T)\ . \tag{7}$$

Cross-Region Income Per Capita Gap Minimization. The cross-region income gap minimization objective was first addressed in Friesz's (1977) multi-objective optimal economic growth model. However, his objective was limited to cross-region income gap minimization but not involving income per capita. In order to balance regional differences in income per capita, develop the lagging regions, and guarantee social stability, it is necessary to reduce the degree of cross-region income per capita gap.

Suppose two regions k and v, $1 \leq k, v \leq n$, and $k \neq v$. The incomes of region k and region v are $Y_k(t)$ and $Y_v(t)$ respectively. Considering population, suppose the populations of region k and region v are $N_k(t)$ and $N_v(t)$; notice the simple labor supply functions of region k and region v are $L_k(t)$, $L_v(t)$, respectively, and $N_k(t) > L_k(t)$, $N_v(t) > L_v(t)$. Then the incomes per capita of region k and region v are $Y_k(t)/N_k(t)$ and $Y_v(t)/N_v(t)$ respectively.

In order to minimize cross-region income per capita gap, it is only necessary to satisfy:

$$Max\ J_2 = -\sum_{k,v=1}^{n} \int_{T_0}^{T} |Y_k(t)/N_k(t) - Y_v(t)/N_v(t)|\, dt\ . \tag{8}$$

There is a unique rule of change of simple labor due to change of investments in every specific sector of specific region according to statistics. These rules change as the level of economic development, human environment, historical and other factors of countries or regions differ. These rules can be simplified as the following homogenous function of degree one:

$$L_{ij}(t) = \lambda_{ij} I_{ij} + C_{ij}, \lambda_{ij} > 0, C_{ij} > 0\ , \tag{9}$$

among which C_{ij} represents the necessary simple labor of sector j in region i, including management, accountants, etc., the number of which will hardly change with investment within a certain scope;λ_{ij} is the labor-investment ratio of sector j in region i, representing changes of simple labor due to changes of investment. Statistic data will be applied to survey labor-investment relationships.

In a word, the output-capital ratio assumption is relaxed, the labor-investment assumption is established, the cross-region cross-section production function is applied to the optimal public investment model, and the optimization objectives

proposed by the predecessors are expanded. There is certain relationship between these objectives. Income maximization and gap minimization are constrained mutually. These objectives must be realized by meeting certain constraints.

3.3 Extension of the Original Constraints

The main contributions of this paper in extending the original constraints lie in modifying taxation and capital transfer loss by dividing the economy into public sectors and private sectors.

Modification of Taxation. Two-sector investment model was first raised by Sakashita (1967). He supposed that in the income $Y_i(t)$ of region i, first $(1 - s_i)$ proportion of the income should be consumed in public sectors and private sectors, and then the rest $s_i Y_i(t)$ should be used to pay taxes at the rate $r(t)$. Therefore $(1 - r(t)s_i Y_i(t))$ should be controlled by private sectors, while $r(t)s_i Y_i(t)$ should be controlled by public sectors, $i = 1, 2$.

Later Friesz (1977, 1978) proposed that the rate of savings for private sectors should be s_i, while the rate of savings for public sectors should be z_i, and these two rates are not identical.

Considering characteristics of taxation, the income $Y_i(t)$ of region i should first be used to pay taxes according to tax rate $r(t)$, ignoring the differences in tax rate between different regions. Suppose all the incomes of public sectors come from taxation, the income amount controlled by the public sectors should be $r(t)Y_i(t)$, applying to consumption and savings of public sectors. The rest $(1 - r(t))Y_i(t)$ should be controlled by private sectors, applying to consumption and savings of private sectors. Suppose all the savings are used to invest, and the rate of savings for public sectors is z_i, the rate of savings for private sectors is s_i. Then public sector investment of region i is $r(t)z_i Y_i(t)$, and total public sector investment of all the regions is $\sum_{i=1}^{n} r(t)z_i Y_i(t)$; similarly, total private sector investment of all the regions is $\sum_{i=1}^{n}(1 - r(t))s_i Y_i(t)$. So the total investment is: $I(t) = \sum_{i=1}^{n}(r(t)z_i + (1 - r(t))s_i)Y_i(t)$. Rewrite it into cross-region cross-sector form: $I(t) = \sum_{i=1}^{n}\sum_{j=1}^{m}(r(t)z_i + (1 - r(t))s_i)Y_{ij}(t)$. According to (2), the time rate of change of capital of all the regions is:

$$K'(t) = \sum_{i=1}^{n}\sum_{j=1}^{m}[(r(t)z_i + (1 - r(t))s_i)Y_{ij}(t)] - \gamma \sum_{i=1}^{n}\sum_{j=1}^{m} K_{ij}(t) \ . \qquad (10)$$

Suppose φ_{ij} is the distribution rate of public sector investment to sector j of region i, ϕ_{ij} is the distribution rate of private sector investment to sector j of region i, $0 \le \phi_{ij}, \varphi_{ij} \le 1$, and $\sum_{i=1}^{n}\sum_{j=1}^{m}\phi_{ij} = 1$, $\sum_{i=1}^{n}\sum_{j=1}^{m}\varphi_{ij} = 1$. Then the time rate of change of capital of region i is: $K_i'(t) = r(t)z_i \sum_{j=1}^{m}\varphi_{ij}Y(t) + (1 - r(t))s_i \sum_{j=1}^{m}\phi_{ij}Y(t) - \gamma \sum_{j=1}^{m} K_{ij}(t)$. So, (10) can be written as:

$$K'(t) = r(t)\sum_{i=1}^{n}[z_i \sum_{j=1}^{m}\varphi_{ij}Y(t)] + (1 - r(t))\sum_{i=1}^{n}[s_i \sum_{j=1}^{m}\phi_{ij}Y(t)] - \gamma \sum_{i=1}^{n}\sum_{j=1}^{m} K_{ij}(t). \quad (11)$$

It should be pointed out that φ_{ij} and ϕ_{ij} are macroeconomic control indexes of great significance. For example, governmental sectors can regulate the emphases of national economy and influence the track of national economy through adjusting the distribution rate of public investment to different sectors of different regions.

Modification of Capital Transfer Loss. The total savings were supposed to be equal to the total investments in the description above. However, Datta-Chandhuri (1967), Domazlicky (1977), and Michel, Pestiean, Thisse (1983) all believed that cross-region capital transfer should cause loss, that is, the investments should be equal to the total available savings.

Suppose the proportion is constant, b_{ik} proportion of the public savings of region i will be lost in the process of capital transfer to region k, $b_{ik} \in [0,1]$, and a_{ik} proportion of the private savings of region i will be lost in the process of capital transfer to region k, $a_{ik} \in [0,1]$, $k = 1, ..., n$. Then the time rate of change of capital of region i is: $K_i'(t) = r(t)(1-\sum_{k=1}^{n} b_{ik})z_i \sum_{j=1}^{m} \varphi_{ij} Y(t) + (1 - r(t))(1 - \sum_{k=1}^{n} a_{ik})s_i \sum_{j=1}^{m} \phi_{ij} Y(t) - \gamma \sum_{j=1}^{m} K_{ij}(t)$. So, (11) can be written as:

$$\begin{array}{l} K'(t) = r(t) \sum_{i=1}^{n} [(1 - \sum_{k=1}^{n} b_{ik}) z_i \sum_{j=1}^{m} \varphi_{ij} Y(t)] \\ +(1 - r(t)) \sum_{i=1}^{n} [(1 - \sum_{k=1}^{n} a_{ik}) s_i \sum_{j=1}^{m} \phi_{ij} Y(t)] - \gamma \sum_{i=1}^{n} \sum_{j=1}^{m} K_{ij}(t) \end{array} \quad (12)$$

3.4 Establishment of Optimal Regional Public Investment Model

To sum up, the double-objective multi-constraint optimal regional public investment model can be expressed as following:

$$Max(J_1, J_2) \quad s.t. \quad (2),\ (4),\ (9),\ (12)\ . \quad (13)$$

The dependent variable of the model is the income of sector j of region i. The control variables of the model are financial tools, including tax rate, weights of every sector of every region, etc. The main external parameters are rate of savings, private sector investment share, discounting rate, and other constraint parameters.

4 Solution and Analysis on the Basis of Genetic Algorithm

The traditional algorithms usually cannot solve optimal economic growth problem with complicated objective functions and constraints well. During the past more than 10 years, GA has been applied to solve all kinds of complicated engineering problems widely, because it is the effective algorithm designed specially for complicated optimization problems developed on the basis of biological simulation, and the powerful stochastic searching optimization technique [12]. As for the optimal public investment model proposed in this paper, because there are many variables, multiple objective functions, and complicated constraints in the model, it should be solved with GA, and in the solving process advantages of GA can be fully shown.

4.1 Handling Objective Functions and Constraints

Resolve ordinary differential equation (2):

$$K(t) = e^{\int(-\gamma)dt}(\int I(t)e^{-\int(-\gamma)dt} + \widetilde{C}) \ . \tag{14}$$

Combine (4), (9) and (14):

$$Y_{ij}(t) = A_{ij}(e^{\int(-\gamma)dt}(\int I(t)e^{-\int(-\gamma)dt} + \widetilde{C}))^{\alpha_{ij}}(\lambda_{ij}I_{ij} + C_{ij})^{\beta_{ij}} \ . \tag{15}$$

Without loss of generality, suppose the government is making public investment plan for period $[T_0, T]$ at T_0. The total public investment is I. At T_0 during $[T_0, T]$ the government plans to invest $\varphi_{ij}I$ in sector j of region i. For given T_0, T, φ_{ij} and I, there are infinite feasible investment paths $I_{ij}(t)$ of sector j of region i. However, most investment process can be divided into two stages in practice: At first a initial $\widetilde{I_{ij}}$ is invested, and the rest is invested gradually. The investment process can be described with the following segmented function:

$$I_{ij}(t) = \begin{cases} \widetilde{I_{ij}} & \text{if } t = T_0 \\ (\varphi_{ij}I - \widetilde{I_{ij}})/(T - T_0) & \text{if } t > T_0 \end{cases} \ . \tag{16}$$

Combine (5), (7), (8), (15) and (16), and it can be seen that both the objective functions and constraints in (13) can be converted into functions of φ_{ij}. Also, some constraints have been incorporated into the objective functions. As a result, (13) can be written as:

$$\begin{array}{rl} Maximize & J_1 = f_1(\varphi_{11}, \varphi_{12}, ..., \varphi_{1m}, \varphi_{21}, ..., \varphi_{2m}, ..., \varphi_{n1}, ..., \varphi_{nm}) \\ & J_2 = f_2(\varphi_{11}, \varphi_{12}, ..., \varphi_{1m}, \varphi_{21}, ..., \varphi_{2m}, ..., \varphi_{n1}, ..., \varphi_{nm}) \\ Subject\ to & f_3 = f_3(\varphi_{11}, \varphi_{12}, ..., \varphi_{1m}, \varphi_{21}, ..., \varphi_{2m}, ..., \varphi_{n1}, ..., \varphi_{nm}) = 0 \\ & f_4 = \sum_{i=1}^{n}\sum_{j=1}^{m}\varphi_{ij} = 1 \end{array} \ , \tag{17}$$

among which f_3 is converted from (12).

When this double-objective programming problem is solved with GA, constraints f_3 and f_4 should be relaxed properly in avoidance of premature convergence of the algorithm. So, let $f_5 = |f_3| \leq \Delta_1$, $f_6 = |f_4 - 1| \leq \Delta_2$, among which Δ_1 and Δ_2 are positive real numbers with very small value. Then (17) can be rewritten as:

$$Max(J_1, J_2) \qquad s.t. \quad f_5 \leq \Delta_1, f_6 \leq \Delta_2 \ . \tag{18}$$

Combine weighted-sum approach and penalty function approach, transform (18) into the following single objective programming model:

$$Max\ J = \pi_1 J_1 + \pi_2 J_2 + F(f_5) + F(f_6) \ , \tag{19}$$

among which $\pi_i \geq 0 (i = 1, 2)$ represents the weight of two optimal objective functions in (19), and $\sum_{i=1}^{2}\pi_i = 1$. $F(f_5)$ and $F(f_6)$ are negative penalty functions which penalize the solutions violate constraints. The design features of

penalty functions are:

$$F(f_5) = \begin{cases} 1.01^g log_G^g, f_5 > \Delta_1 \\ f_5, \quad f_5 \leq \Delta_1 \end{cases} \text{ and } F(f_6) = \begin{cases} 1.01^g log_G^g, f_6 > \Delta_2 \\ f_6, \quad f_6 \leq \Delta_2 \end{cases},$$

among which g is the number of evolution generations, and G is the maximal number of generations (see [13]).

4.2 Encoding and Decoding Approaches

It is a key to GA to encode and decode the solutions into chromosomes. In the optimal cross-region public investment model, a binary string [14] of 7 bits can be used to represent φ_{ij}. The value of the binary string is a percent. If it exceeds 100%, the value is defined as 100%. For example, if the corresponding binary string of φ_{11} in the chromosome is 0011001, so $\varphi_{11} = 15\%$. Then after encoding $m \times n$ variables, the chromosomes of optimal cross-region public investment model are binary strings with the length of $7 \times m \times n$, i.e. $A[1..7 \times m \times n]$. Let array element $Solution[i][j]$ store φ_{ij}, and the decoding algorithm of solution to optimal cross-region public investment model is as following: INPUT: $A[1..7 \times m \times n]$; OUTPUT: $Solution[1..i][1..j]$.

```
Algorithm: Decoding BEGIN
    INTEGER i:=1; INTEGER j:=1; INTEGER k:=1; FLOAT temp:=0.0;
    WHILE (i<=n) DO
        {WHILE (j<=m) DO
            WHILE (k<=7) DO
                {temp:=temp+A[7*(i-1)*m+7*(j-1)+k]*2^(7-k);
                 k:=k+1;}
             IF (temp/100 > 100%)
                temp=100%;
             Solution[i][j]:=temp; j:=j+1; k:=1; temp:=0.0;}
        i:=i+1; j:=1;}END
```

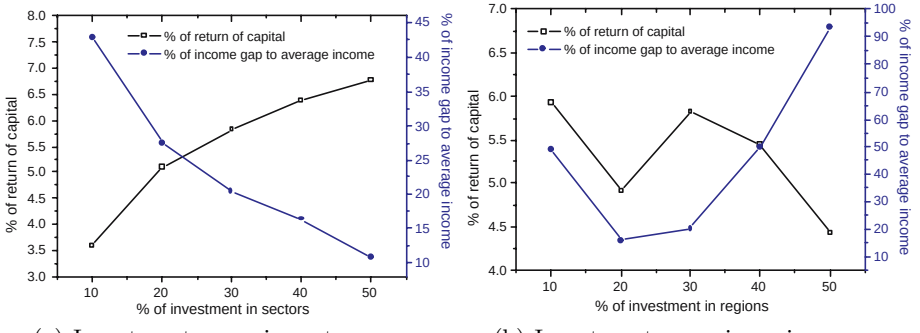
4.3 Case Study and Result Analysis

A case study is carried out to verify effectiveness of the optimal cross-region public investment model. Values are assigned to relevant constant parameters in the model first, and weights of the three optimal objective functions are adjusted dynamically to record the changing courses of two optimal objective functions secondly. As a result, it can be concluded whether the public investment decision on basis of the proposed GA-based model is able to adjust economic growth rate and control income gap.

Suppose there are two regions with two sectors each. The relevant parameters are: $A = 5, \alpha_{ij} = 0.6, \beta_{ij} = 0.4, s_i = 50\%, z_i = 90\%, \xi_{11} = 0.6, \xi_{12} = 0.4, \xi_{21} = 0.7, \xi_{22} = 0.3, \omega_1 = 0.4, \omega_2 = 0.6, \phi_{ij} = 0.25, r(t) = 0.3, a_{ij} = b_{ij} = 0.1, \gamma = 0.1, \lambda_{11} = 1, \lambda_{12} = 0.5, \lambda_{21} = 2, \lambda_{22} = 0.2, C_{11} = 1, C_{12} = 2, C_{21} = 1, C_{22} = 2, N_1 = 15, N_2 = 20$ (in thousands), $Y_{11}(0) = Y_{12}(0) = Y_{21}(0) = Y_{22}(0) = 0, [T_0, T] = [0, 1], I(0) = RMB200$ (in millions).

The optimal public investment model is resolved with GA according to the parameters above. Chromosome encoding and decoding approaches are shown in Subsection 4.2. The chromosomes are selected based on roulette wheel selection approach for next population. Two-point crossover and uniform mutation is performed. And the population is evaluated based on fitness function (19). Then we get $\varphi_{ij} = 0.25$. Change the relevant parameters, the distribution proportions of φ_{ij} will change correspondingly. Results are demonstrated in Fig. 1.

In Fig. 1(a), the investment is averagely divided between regions. The return of capital increases and the percentage of income gap to average regional income decreases as the investment to the more labor-intensive sector increases.



(a) Investments vary in sectors (b) Investments vary in regions

Fig. 1. Investments vary in sectors and regions

In Fig. 1(b), the investment is averagely divided between sectors. The return of capital fluctuates but tends to decrease as the investment to the less-developed region increases. And the percentage of income gap to average regional income fluctuates but tends to increase in the same process as above.

As can be seen from Fig. 1, different proportions of public investment may cause economic growth and income gap to vary; proportions of public investment may be used for macro-control of different regions and sections; and GA-based solutions to the model can provide feasible foundation for public investment decision.

5 Conclusions

To sum up, the cross-region cross-sector Cobb-Douglas aggregative production function is established first in this paper. Secondly two objective functions involving both efficiency and equity and such constraints as taxation and capital transfer loss are combined to form a double-objective, multi-constraint, cross-region, cross-sector optimal regional public investment model. Thirdly GA is applied to solve the model. And finally a case study validates the effectiveness of both the model and GA-based solution.

References

1. Gaube, T.: Public Investment and Income Taxation: Redistribution vs. Productive Performance. Journal of Economics, Vol. 86 (2005) 1-18
2. Basu, P., Bandyopadhyay, D.: What Drives the Cross-country Growth and Inequality Correlation. Canadian Journal of Economics, Vol. 38, No. 4 (2005) 1272-1297
3. Brecher, R.A., Chen, Z.Q., Choudhri, E.U.: Dynamic Stability in a Two-country Model of Optimal Growth and International Trade. Journal of Economic Dynamics & Control, Vol. 29 (2005) 583-594
4. Van, C.L., Vailakis, Y.: Recursive utility and optimal growth with bounded or unbounded returns. Journal of Economic Theory, Vol. 123 (2005) 187-209
5. Jouini, E., Napp, C.: Convergence of Utility Functions and Convergence of Optimal Strategies. Finance and Stochastics, No. 8 (2004) 133-144
6. Aurell, E., Muratore-Ginanneschi, P.: Growth-Optimal Strategies with Quadratic Friction over Finite-Time Investment Horizons. International Journal of Theoretical and Applied Finance, Vol. 7, No. 5 (2004) 645-657
7. Saglam, H.C.: Optimal Growth Models and the Lagrange Multiplier. Journal of Mathematical Economics, Vol. 40 (2004) 393-410
8. Gómez, M.A.: Optimal Fiscal Policy in the Uzawa-Lucas Model with Externalities. Economic Theory, No. 22 (2003) 917-925
9. Wu, J.H.: Optimal Regional Growth and Public Investment Models: Theory and Application to the Efficiency-Equity Issue in Taiwan. PQDD doctoral dissertation, Dissertation Abstracts International, Vol: 48-04, Section A (1987) 1035
10. Ramsey, F.: A mathematical theory of saving. Economic Journal, Vol. 38 (1928) 543-559
11. Romer, D.: Advanced Macroeconomics. 2nd Edn, McGraw-Hill Co., Inc. (2000)
12. Gen, M., Cheng, R.W.: Genetic Algorithms and Engineering Optimization. John Wiley & Sons, Inc. (2000)
13. Pan, Y., Yu, Z.W., Liu, K.J., Dou, W.: A New Multi-Objective Programming Model of QoS-based Multicast Routing Problem. Computer Engineering and Application, No. 19 (2003) 155C157
14. Guo, H.Y., Zhang, L., Jiang, J.: Two-Stage Structural Damage Detection Method with Genetic Algorithms. Journal of Xi'an Jiaotong University, Vol. 39, No. 5 (2005) 485-489

Multi-population Genetic Algorithm for Feature Selection

Huming Zhu[1], Licheng Jiao[1], and Jin Pan[2]

[1] Institute of Intelligent Information Processing, Xidian University, Xi'an, Shaan'xi 710071, China
{Zhuhum, lcjiao}@mail.xidian.edu.cn
[2] Dept. of Computer and Information Eng, Xi'an communication institute, Xi'an, Shaan'xi 710106, China
Panjin_163@163.com

Abstract. This paper describes the application of a multi-population genetic algorithm to the selection of feature subsets for classification problems. The multi-population genetic algorithm based on the independent evolution of different subpopulations is to prevent premature convergence of each subpopulation by migration. Experimental results with UCI standard data sets show that multi-population genetic algorithm outperforms simple genetic algorithm.

1 Introduction

Dimensionality reduction is a common preprocessing step used for pattern recognition and classification applications. Nowadays, many Dimensionality Reduction Methods are introduced, as feature selection, sparse representations, feature construction, etc [2] [4] [5]. However, feature selection has received considerable attention. As we know, Feature selection is the problem of selecting a subset of d features from a set of D features based on some optimization criterions. The primary purpose of feature selection is to design a more compact classifier with as little performance degradation as possible in which the features removed should be useless, redundant, or of the least possible use.

Genetic algorithms (GAs) are heuristic methods for solving computationally difficult problems using biologically inspired notions of Darwinian evolution [3]. They have been applied to a variety of problems, from function optimization to feature selection. GAs frequently have an advantage over many traditional local search heuristic methods when search spaces are highly modal, discontinuous, or highly constrained. Several authors have explored the use of genetic algorithms for feature subset selection for neural network and nearest neighbor classifiers, but the primary complaints toward genetic algorithms are the occurrence of premature convergence [6],[14]. Premature convergence can be subdued through variations of operations such as crossover, selection, mutation, and through alterations of parameters such as population size, crossover rate, and mutation rate. Much research has been done in developing improved genetic algorithms. Past research has focused on the improvement of operators and parameter settings and indicates that premature convergence is still the preeminent problem in GA's.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 480 – 487, 2006.

This paper describes the application of a multi-population genetic algorithm (MGA) to the selection of feature subsets for classification. MGA is an algorithm whose architecture is specifically designed to confront the causes of premature convergence. The communication between sub-populations is assured by a migration process: after some generations several elements leave their population and migrate to another. This process has in important role in preserving the population diversity, thus in avoiding premature convergence cases. Experimental results with UCI standard data sets show that multi-population genetic algorithm outperforms simple genetic algorithm.

The paper is organized as follows. The next section is a brief introduction to feature selection and nearest neighbor classification. Section 3 proposes a multi-population genetic algorithm for feature selection (MGAFS). Our experimental results and a discussion are presented in Section 4. Finally, Section 5 concludes the paper.

2 Feature Selection and Nearest Neighbor Classification

Feature selection is a common and key problem in many classification and regression tasks. It can be viewed as a multi-objective optimization problem, since, in the simplest case, it involves feature subset size minimization and performance maximization.

2.1 Feature Selection Problem

The problem of feature selection is defined as follows: given a set of candidate features, select a subset that performs the best under some classification system. This procedure can reduce not only the cost of recognition by reducing the number of features that need to be collected, but in some cases it can also provide a better classification accuracy due to finite sample size effects [4].A number of approaches to feature subset selection have been proposed in [9],[10],[15]. These approaches involve searching for an optimal subset of features based on some criteria of interest. Since exhaustive search over all possible subsets of a feature set is not computationally feasible in practice ,a number of authors have explored the use of GA for feature subset selection[2],[12],[13]. Kudo and Sklansky [5] have compared several algorithms for feature selection and concluded that GAs are suitable when dealing with large-scale feature selection.

Feature subset selection algorithms can also be classified into two categories based on whether or not feature selection is done independently of the learning algorithm used to construct the classifier. If feature selection is performed independently of the learning algorithm, the technique is said to follow a filter approach, Otherwise it is said to follow a wrapper approach.

2.2 Nearest Neighbor Classification

The nearest neighbor (1-NN) classifier is one of the oldest and simplest methods for performing general, non-parametric classification. It can be represented by the following rule: to classify an unknown pattern, choose the class of the nearest example in the training set as measured by a distance metric. A common extension is to choose the most common class in the k nearest neighbors (kNN). Despite its simplicity, the nearest neighbor classifier has many advantages over other methods.

For example, it can learn from a small set of examples, can incrementally add new information at runtime, and can give competitive performance with more modern methods such as decision trees or neural networks.

We propose a multi-population GA-Based Wrapper Feature Selection for nearest neighbor classification.

3 Multi-population Genetic Algorithm for Feature Selection(MGAFS)

Genetic algorithms are able to escape from local optima by means of the crossover and mutation operators, and to explore a wide range of search space when the selection pressure is properly controlled. However, the primary complaint toward simple genetic algorithms (SGA) is the occurrence of premature convergence. To improve the fine-tuning capability of SGA, multi-population GAs have been developed in many applications, including function optimization [1], the 0/1 knapsack problem [7], and the file allocation problem [11]. Multi-population genetic algorithm is an extension of traditional single-population genetic algorithms by dividing a population into several isolated sub-populations within which the evolution proceeds and individuals are allowed to migrate from one sub-population to another. Generally, There areãsubpopulations. Individual in MGA are migrated after every ù(migrate interval) generations and the best-worst migration policy is used. That is, the best ñ(migrate rate)of individuals in one subpopulation are selected to migrate its neighbor subpopulations, and replace the worst individuals therein.

In recent years, MGA have been recognized as being more effective both in speed and solution quality than single-population genetic algorithms.

We propose a multi-population GA for the feature selection problem. It is outlined below.

A Multi-population genetic algorithm can be described as Figure 1.

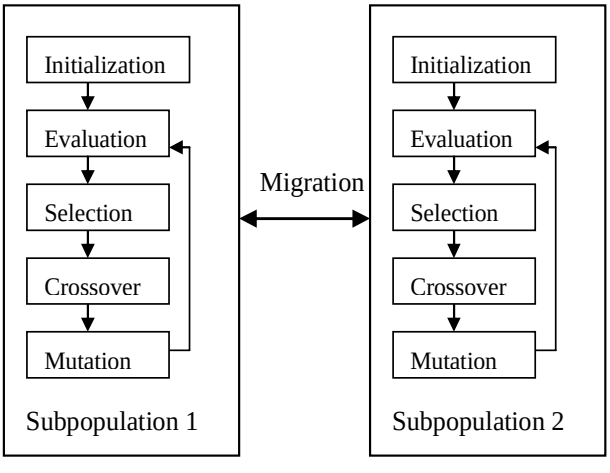


Fig. 1. A typical MGA with two subpopulations

3.1 Chromosome Encoding

For the feature selection problem, each individual in the population represents a candidate solution to the feature subset selection problem [13]. Let D be the total number of features available to choose from to represent the patterns to be classified. It is represented by a binary vector of dimension D (where D is the total number of features). If a bit is a 1, it means that the corresponding feature is selected. A value of 0 indicates that the corresponding feature is not selected. As an example, chromosome 0100010 means that the second and sixth features are selected.

The fitness of an individual is determined by evaluating the Nearest Neighbor classifier using a training set whose patterns are represented using only the selected subset of features.

3.2 Initial Population

Instead of generating a single population as in SGA, MGAFS uses two subpopulations: subpopulation 1 and subpopulation 2.

The generation of the initial subpopulation is straightforward, as shown below. The function rand (subpop_size,D) generates a random floating-point array that the value of the elements contained in the array is within [0,1]. The expected number of selected features in an arbitrary initial solution is d. Initial sub-population:

for (i = 1 to subpop_size)
for (j = 1 to D)
if (rand (i,j) < d/D) g =1; else g= 0; g represents jth gene in ith chromosome

MGAFS initially creates subpopulation 1 and subpopulation 2. To enhance exploration early, each initial subpopulation is generated with a predetermined bias. One subpopulation is biased to more 0's in individual chromosomes, the other is biased to more 1's,

The initial subpopulation 1 is randomly created with a bias of 0.4.The initial subpopulation 2 is created with a bias of 0.6 giving an approximate initial ratio of four 0's to six 1's.

3.3 Fitness Function

The fitness function is straightforward since each individual in the population represents a selected feature subset, x, and the fitness function is simple.

Fitness(x) = accuracy(x)

Where fitness(x) is the fitness of the feature subset represented by x, accuracy(x) is the test accuracy of the Nearest Neighbor classifier trained using the feature subset represented by x.

3.4 Selection

Our experiments were run using a genetic algorithm using rank-based selection strategy. The probability of selection of the highest ranked individual is p (where p is

a user-specified parameter), that of the second highest ranked individual is p (1- p), that of the third highest ranked individual is p $(1- p)^2$; that of the last ranked individual is 1-(sum of the probabilities of selection of all the other individuals).

3.5 Migration

MGA is more complex than simple genetic algorithms. In particular, the migration of individuals from one subpopulation to another is controlled by several parameters like [1]: (1) the topology that defines the connections between the subpopulations, (2) a migration rate that controls how many individuals migrate, (3) a migration interval that affects the frequency of migrations, and(4) migration policy defines what is being sent.

In the experiments we used a model with two identical subpopulations (the same parameters) evolving synchronously and a best-worst migration policy, in which the best individual is chosen from the source subpopulation and the worst individual is replaced in the target population occurred after everyùgenerations. We varied the migration interval from 1 to 100.

3.6 Parameters

No systematic parameter optimization process has so far been attempted, but the following parameter set was used in our experiments. Tuning the values to be suitable for a specific data set may give rise to improved performance.

Parameter setting:

Control procedure: generational procedure
Subpopulation size = 10
pc (crossover probability) = 1.0
pm (mutation rate) = 0.1
q (in rank-based selection) = 0.25
T (maximum generation) =100
Number of experiments: 10 runs (get the average)
The parameter settings were based on results of several preliminary runs.

4 Results and Discussion

4.1 Dataset

The two data sets used to test the algorithms are summarized in Table 1. Two data sets were chosen from the UCI repository [8] with the condition that the number of features be 30 or greater, that there be no missing feature, and that all of the features be numeric.

Ionosphere. This data set consists of 351 patterns, with 34 attributes and one output with two classes, good or bad, with good implying evidence of some type of structure in the ionosphere and bad the lack of such evidence.

Sonar. (Mines vs. Rocks) The task is to discriminate between sonar signals bounced off a metal cylinder and those bounced off a roughly cylindrical rock. The dataset has 208 cases, 60 inputs and one output of two classes.

Table 1. Data Sets Used for Our Experiment

dataset	Number of samples	Number of features	Number of classes	evaluation	
				classifier	Leave one out
Ionosphere	351	34	2	1-NN	yes
Sonar	208	60	2	1-NN	yes

4.2 Classification Accuracy

Table 2 summarizes the performance of the three feature selection algorithms for the two data sets. SGA-P represents proportional selection; SGA-R represents rank-based selection .X and y in x(y) represent the maximum and average recognition rates.

The evaluation was conducted in terms of the criteria: classification accuracy. Accuracy was measured by determining the recognition rate of 1-NN classifiers without rejection.

Through the experimental results obtained using the two data sets, the performances of our MGAFS and simple genetic algorithms are compared, in terms of the accuracy of their classification. We were able to conclude that our MGAFS is superior to the SGA. However, the average quality of the final population was much lower in the MGAFS. It is trade off between exploration and exploitation.

Table 2. Recognition Rates for the two Data Sets (Unit: %)

dataset	All features	SGA-P	SGA-R	MGAFS
Ionosphere	86.61	92.88(87.90)	94.59(93.21)	95.16(91.74)
Sonar	82.69	93.75(84.74)	94.23(91.37)	95.19(90.7)

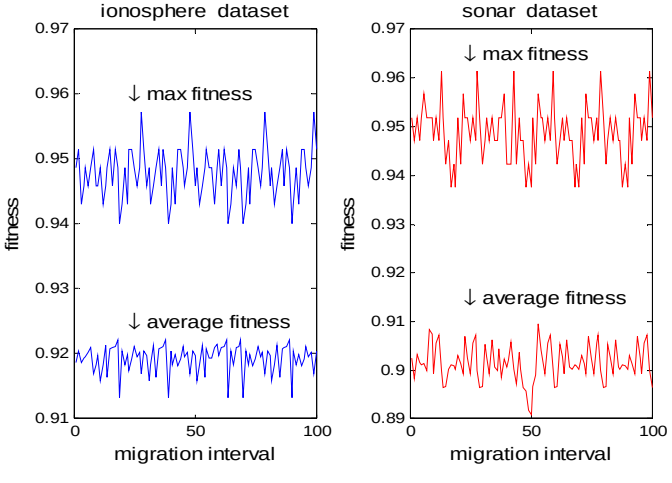


Fig. 1. Average fitness and max fitness in runs with different migration intervals

Fig.1 indicates that migration frequency has a significant effect on the quality of the solutions. Migration can prevent premature convergence. There is an interesting question being raised here: when is the right time to migrate? It seems that there exists a relationship between migration frequency and the appearance of the max fitness; however, more research is necessary to understand the effect of migration on the quality of the search in MGAFS.

5 Summary and Discussion

An approach to feature subset selection using a multi-population genetic algorithm for the nearest neighbor (1-NN) classifiers is proposed in this paper. The results presented in this paper indicate that MGAFS offer an attractive approach to solving the feature subset selection problem in nearest neighbor pattern classifiers. However, additional experiments with other datasets and varying subpopulation sizes and migrations sizes are needed to confirm and extend the results. We plan to implement parallel multi-population genetic algorithm in cluster computers.

References

1. Cantú-Paz, E.: A survey of parallel genetic algorithms. Reseaux et Systems Repartis 10(1998)141-171
2. Dash, M., Liu, H.: Feature Selection for Classification. Intelligent Data Analysis 1 (1997) 131-156
3. Davis, L.: Handbook of Genetic Algorithms. Van Nostrand Reinhold, New York (1991)
4. Jain, A., Zongkerm, D.: Feature selection: Evaluation, application, and small sample performance. IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (1997) 153-158
5. Kudo, M., Sklansky, K.: Comparison of algorithms that select features for pattern classifiers. Pattern Recognition 33 (2000) 25–41
6. Kuncheva, L.I., Jain, L.C. :Nearest Neighbor Classifier: Simultaneous Editing and Feature Selection. Pattern Recognition Letters 20(1999) 1149-1156
7. Wen-Yang, Lin., Tzung-Pei, Hong., Shu-Min, Liu.: On adapting migration parameters for multi-population genetic algorithms. In: Proceedings of IEEE International Conference on Systems, Man, Cybernetics(2004) 5731-5735
8. Murphy, P. M., Aha, D. W.: UCI Repository of machine learning databases. University of California, Irivne, Department of Information and Computation Science (1996)
9. Narendra, P.M., Fukunaga, K.: A Branch and Bound Algorithm for Feature Subset Selection, IEEE Trans. Computers 26(1977) 917-922
10. Pudil, P., Ferri, F. J., Novovičová, J., Kittler, J.: Floating Search Methods in Feature Selection. Pattern Recognition Letters (1994) 1119- 1125
11. Potts, J. C., Giddens, T.D., Yadav, S.B.: The Development and Evaluation of an Improved Genetic Algorithm Based on Migration and Artificial Selection. IEEE Transactions on Systems, Man, and Cybernetics 24 (1994) 73-86
12. Raymer, M.L., Punch, W.F., Goodman, E.D., Kuhn, L.A., Jain, A.K.: Dimensionality Reduction Using Genetic Algorithms. IEEE Trans. Evolutionary Computation 4(2000) 164-171

13. Siedlecki, W. , Sklansky, J.: A note on genetic algorithms for large-scale feature selection. Pattern Recognition Letters 10 (1989) 335-347
14. Jihoon, Yang., Honavar, V.: Feature Subset Selection Using a Genetic Algorithm. IEEE Intelligent Systems 13(1998) 44-49
15. Bin, Yu., Baozong, Yuan.: A More Efficient Branch and Bound Algorithm for Feature Selection. Pattern Recognition 26 (1993)883-889

Using Wearable Sensor and NMF Algorithm to Realize Ambulatory Fall Detection

Tong Zhang[1], Jue Wang[1], Liang Xu[2], and Ping Liu[1]

[1] Key Laboratory of Biomedical Information Engineering of Education Ministry, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, P.R. China
zhangtong3000@163.com, juewang@mail.xjtu.edu.cn, liuhuangs@163.com
[2] School of Computer Science and Technology of Xidian University, Xi'an, Shaanxi 710071, P.R. China
xuliang5110@mail.xidian.edu.cn

Abstract. Falls in the elderly people often cause serious physical injury, result in fracture, cerebral haemorrhage, even death. To find falls as earlier as possible is very important to rescue the subjects and facilitate the rehabilitation in the future. In this paper, we use a wearable tri-axial accelerometer to monitor the movement parameters of human body, and propose a novel fall detection algorithm based on non-negative matrix factorization (NMF). The input vectors are the acceleration sequences of the transverse section and the vertical axial of human body, and these vectors are decomposed via NMF. And then, a k-nearest neighbor method is applied to determine whether a fall occurred. The results show that this method can detect the falls effectively.

1 Introduction

The fall is very dangerous to the elderly people, often causes serious results as fracture**,** cerebral haemorrhage, even death. To find the fall as earlier as possible is highly important to rescue the subjects and facilitate the rehabilitation in the future.

There are basically three ways to realize fall detection[1]: video based, acoustic based and wearable sensor based. Because video and acoustic devices (e.g. camera and microphone) are generally installed in fixed position, the wearable sensor based system is more suitable for ambulatory monitoring, but currently, there exist many problems in the corresponding algorithms, involve lacking of adaptability, deficiently in classification precision, etc. Hence, we try to find some intelligent methods to solve these problems. This paper uses a tri-axial accelerometer to capture the movement signals of human body and proposes a novel algorithm based on NMF to detect falls. The input vectors are the acceleration sequences of the transverse section and the vertical axial of human body, these vectors are decomposed via NMF, and then a k-nearest neighbor method is applied to realize the classification. We got approximately 600 examples for training and testing, and the results show that this method can detect the falls effectively.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 488 – 491, 2006.

2 Methods

2.1 Fall Analysis and the Input Vectors

We attached the sensor "MMA7260Q", a tri-axial low-g micro-machined resistance accelerometer with the outputs of a_x, a_y, and a_z, to a belt and bound it to abdomen (between the lumbar vertebrae L2-L4). If the volunteer standing upright, the axis y direct upwards (against the vertical axis), the axis x parallel to frontal axis, the axis z coincide with sagittal axis. We denote a_s as the vector sum of $a_x + a_z$, a_{totle} as $a_s + a_y$, and therefore a_s indicates the acceleration in the transverse section of human body.

The fall is a non-cyclical movement, there are basically three steps in turn: daily activities, falling and a period of motionless (or only a little motion). The motionless interval will last for several seconds or more with a_{totle} nears to g; the falling will generally be completed within 0.4-1.0 second with acute varieties in a_s and/or a_y. So we detect $|a_{totle}|$ at first, if it nears $|g|$ in one second, we will consider the subject in motionless, then we backdate the data to find a 1.5s-length-section with acute changes in $|a_s|$ and/or a_y. The sample rate is 512 points per second, and after low-pass filter, we get the 768*2 sequence as an input vector. Because a_y has negative values, we plus a positive constant of $5g$ to ensure the input vectors be non-negtivity.

2.2 NMF Algorithm

Non-negative matrix factorization (NMF) is a method for space decomposition and feature extraction, it factorizes an input matrix into a basis vector matrix and a project matrix with non-negtivity constrains. Each input vector can be expressed as a linear combination of the basis vectors, both the basis vector matrix and the project matrix are sparse [2].

Given an input martix with $m \times n$ dimensions, denote as V_{mn}, where $V_{iu} \in R$ and $V_{iu} \geq 0$, then NMF constructs the following factorization:

$$V_{iu} \approx (WH)_{iu} = \sum_{\alpha=1}^{r} W_{i\alpha} H_{\alpha u} \ , \ W_{i\alpha} \geq 0, H_{\alpha u} \geq 0 \tag{1}$$

When the evaluation function is defined by Kullback-Leibler divergence, the equation (1) becomes an optimization problem as follows:

$$\min_{W,H} \left(\sum_i \sum_j \left(V_{ij} \log \frac{V_{ij}}{(WH)_{ij}} - V_{ij} + (WH)_{ij} \right) \right) \tag{2}$$

s.t.

$$W_{i\alpha} \geq 0, H_{\alpha u} \geq 0 \text{ , and } \sum_i W_{i\alpha} = 1, \forall \alpha$$

the local minimum of equation (2) can be found via the following iterative rules[2]:

$$W_{i\alpha} \leftarrow W_{i\alpha} \sum_j \left(\frac{V_{ij}}{(WH)_{ij}} \frac{H_{\alpha j}}{\sum_k H_{\alpha k}} \right) , \ W_{i\alpha} \leftarrow \frac{W_{i\alpha}}{\sum_j W_{j\alpha}} , \ H_{\alpha u} \leftarrow H_{\alpha u} \sum_j \left(W_{j\alpha} \frac{V_{ju}}{(WH)_{ju}} \right) \tag{3}$$

2.3 Classification

All available data are random divided into two subsets, for training and testing respectively. The training set V_{mn} , includes positive (fall) and negative (daily activity) samples with approximately the same number, is decomposed via NMF. After the decomposition, every row vector of H is normalized, then the k-nearest neighbor algorithm is applied for classification, and a threshold T_d is introduced to avoid a wild point be classified in positive sample. When a test vector x is factorized and the output project vector h_t has k nearest neighbors in $\{h_j\}$, the decision function will be:

$$if \ \ (k_p > k_q, k_p + k_q = k) \ \text{and} \ \ (d_{\min}(h_t, h_{kp}) < T_d) \tag{4}$$

$$then \qquad x \in \omega_p \qquad , \qquad otherwise \qquad x \in \omega_q$$

where $\{h_j\}$ is the row vector set of H, ω_p is the positive sample class, ω_q is the negative sample class. To the nearest neighbors, k_p is the number of the positive samples, k_q is the number of the negative samples, and h_{kp} expresses arbitrary the positive samples, d_{min} indicates the minimum distance. The value of r and k are determined by global search, where we let $r \in \{20, 25, \ldots, 100\}$ and $k \in \{11, 13, \ldots, 51\}$, and we decompose V_{mn} 16 times for each (r, k) because the decomposition is not unique. Then we select the optimal (r, k, W) based on the empirical error minimization. $T_d = pd_{np}$, where d_{np} indicates the minimum distance between the positive and negative samples in the training set, and p is a positive coefficient.

3 Results and Discussion

We arranged the experiments on 6 categories, as shown in table.1, where the 3rd category are the movements of groveling on the ground quickly, sat down heavily, etc.; the 4th is similar to the 3rd, but the movements were slower; the 5th involves walking, jogging, etc.; and the 6th means some acute daily activities, e.g. jump and gymnastics. Our aim is to detect fall, so the negative samples (daily activities) we selected are basically more similar to falls than other daily activities (e.g. cooking, sleeping). And in order to reduce the risk in experiments, the elderly people only attended category 4 and 5, the category 2 were implemented with a dummy.

There were 602 examples obtained, 331 for training and 241 for testing. After trial, we got r=50, k=21,and T_d =0.17. The examination results are shown in table.1. For category 1, 2, 4 and 5, the ratio of correctness r_c is high, indicates that the algorithm has a well performance on distinguishing falls from non-acutely daily activities; for category 3 and 6, r_c does not exceed 85%, that shows the capability of our method is not well in differentiating the real falls from the acutely movement. But fortunately, the activities corresponding to category 3 and 6 are rarely to the elderly people.

Almost all the current algorithms have a connotative condition, i.e., the tilt degree of human body will be changed significantly when a fall occurred, but that's not always true, e.g. a person slide down and sitting on the ground, the tilt degree will be changed slightly. Our algorithm has no this implied condition, it is more adaptive and

efficient. We also developed a fall detection method based on 1-class support vector machine, it has similar performance to the NMF based algorithm, but the pre-processing is more complex than the latter.

Table 1. The examination results

category	correct	incorrect	ratio of correctness(%)
1. fall down on soft cushion	63	1	98.4
2. fall down on hard floor, stair, etc	86	4	95.6
3. critical movements	15	3	83.3
4. sub-critical movements	28	0	100.0
5. low-intensity daily activities	28	1	96.6
6. high-intensity daily activities	9	3	75.0
total	229	12	95.0

4 Conclusion and Future Works

Capturing the movement parameters of human body by wearable sensors, and using NMF algorithm, we can realize the ambulatory fall detection credibly and efficiently. In the future, we intend to extend this method to realize the monitoring to the patients with some chronic diseases (e.g. Parkinson's disease).

Acknowledgments

The authors acknowledge the support from National Natural Science Foundation of China, grant 60271025.

References

1. Luo, S., Hu, Q.: A Dynamic Motion Pattern Analysis Approach to Fall Detection. IEEE International Workshop on Biomedical Circuits&Systems, Singapore (2004) S2.1_5-S2.1_8
2. Lee, D. D. and Seung, H.S.: Algorithms for Non-negative Matrix Factorization. The 14th Annual Conference on Neural Information Processing Systems, Vol. 13, (2000) 556-562

Actor Based Video Indexing and Retrieval Using Visual Information

Mohammad Khairul Islam, Soon-Tak Lee, and Joong-Hwan Baek

Department of Information and Telecommunications Engineering, Hankuk Aviation University, 200-1, Hwajeon-dong, Goyang-city, Gyeonggi-do, 412-791, Korea
mki@hau.ac.kr, soontaklee@yahoo.co.kr, jhbaek@hau.ac.kr

Abstract. Content-based video indexing and retrieval algorithms are presented in this paper that aim at temporally indexing a video sequence according to actors. Our system splits a video into a sequence of a few representative frames. We use color information and then SGLD matrix on the representative frames for face region detection. Detected faces are used to build a face database. We construct eigen faces applying PCA on the faces in the face database for extracting important features. Extracted features are then used in MPM for identifying the input face from the training faces. Experimental result shows that our approach can correctly recognize 95.3% and 90.84% of the faces from the AT&T face database and video sequence respectively.

1 Introduction

Multimedia retrieval and browsing techniques improve a service quality of the multimedia and value of contents that the media provider possesses, and it is a salient research topic in the multimedia service industry. Also, the more multimedia information is utilized in the various field of industry including internet based e-library, digital studio and VOD (video on demand) etc., the more techniques for video retrieval and database management will be demanded increasingly. So, research for multimedia indexing and retrieval has to be worked continuously. Its technique must be diffused by the production on commercial scale and new techniques must be developed for an efficient multimedia indexing and retrieval. There is a recent research of semantic based video analysis. In the semantic video indexing, adjacent shots are clustered and clustered shots compose the story units [2]. Video summarization based story units provides a higher-level video context, however, not every shots contains a meaningful thematic topic [1]. For that reason, specific event detection methods are demanded for a higher semantic level so as to better reveal, represent and abstract the video content. The typical research for content-based video indexing includes face detection [3], speaker identification and character recognition [4].

In this paper, we propose an efficient face region detection and identification algorithm using visual information for character recognition in future. The paper is organized as follows. Section 2 demonstrates the framework of the approach. Section 3

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 492 – 501, 2006.

represents the skin color detection, along with statistical attribution for skin color. Section 4 explains SGLD (spatial gray level dependence) matrix as textural feature value and describes classification tree of CART (classification and regression tree). Section 5 and 6 describes PCA (Principal Component Analysis) and MPM (Minimax Probability Machine) respectively. In Section 7, an experimental result shows that the proposed method has a high performance. Finally, the conclusion and discussion of future work are given in Section 8.

2 Framework for Our Approach

Given an arbitrary image, the goal of face identification is to determine whether or not there are any faces in the image, if present, return the image location and extent of each face, and finally identify the face from face database. In our work, we go for actor faces in video frames. So, for key frame selection, we build skin color statistics at first from known face region of an actor, and apply Bayesian decision rule to input frames. For face region detection, we segment input image, extract feature values, and classify those values. Using detected faces we look for actor faces in the face database for identification. So, our approach works as like as the following diagram.

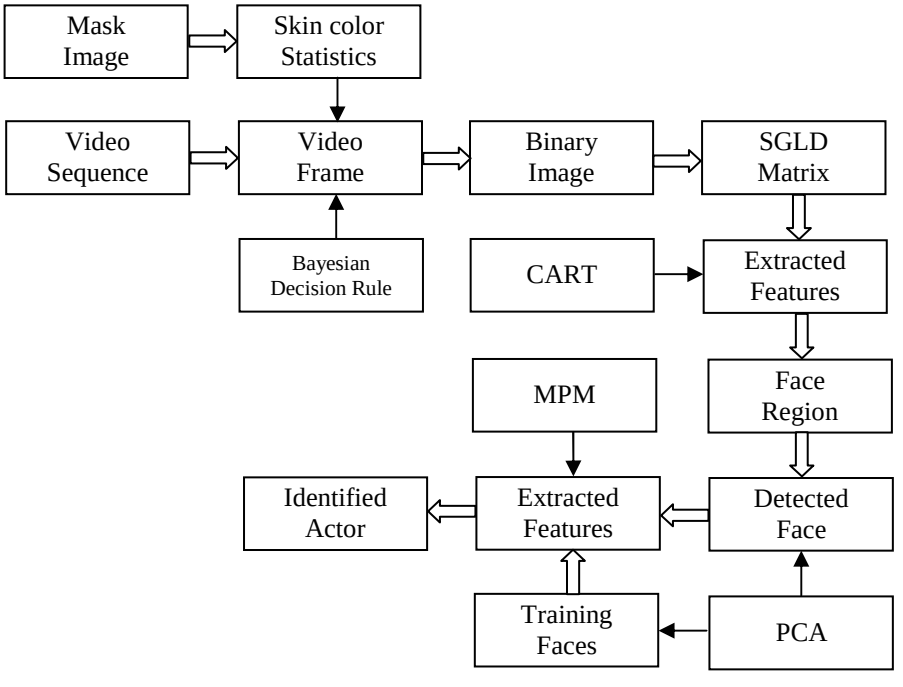


Fig. 1. General schema of our approach to video indexing. In the figure, the symbol ⟶ stands for "applied to" and ⇨ for "results".

3 Generation of Face Candidates Using Skin Color

Human skin color has been used and proven to be an effective feature in many applications from face detection to hand tracking [5]. Although different people have different skin color, several studies have shown that the major difference lies largely between their luminance rather than their chrominance [1]. Out of various color spaces like RGB, HIS, YCbCr, YIQ, and CIE, we use YCbCr color space for face candidate.

3.1 Building Skin Color Statistics

The conversion between RGB color space and YCbCr color space is given as following Equation (1). RGB values normalized to [0, 1] are transformed into [0, 1] luminance Y and [-0.5, 0.5] chrominance Cb, Cr.

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} \bullet \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

We manually generate mask image to extract skin color from training images sample. The color of the mask region is used as training samples for the skin and other for non-skin class. Most skin pixels are distributed like the two dimensional Gaussian density function with a negative mean for Cb, positive mean for Cr, and low variation. So, we assumed that the probability distribution function of skin color pixels is the bivariate Gaussian distribution illustrated in equation (2). The parameters μ, σ and ρ denote the mean, standard deviation and correlation coefficient, respectively.

$$p(x,y) = \frac{e^{-\frac{1}{2(1-\rho^2_{xy})}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - \frac{2\rho_{xy}(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2_{xy}}} \tag{2}$$

3.2 Bayesian Decision Rule

When there are continuous d dimensional features, and their ranges overlap for the two classes, optimal decision boundary between two classes, C_i and C_j, will be where their posterior probabilities are equal like equation (3).

$$P(C_i \mid \mathbf{x}) = P(C_j \mid \mathbf{x}) \tag{3}$$

$$P(C_i)\, p(\mathbf{x} \mid C_i) = P(C_j)\, p(\mathbf{x} \mid C_j) \tag{4}$$

$$D = P(C_i)\, p(\mathbf{x} \mid C_i) - P(C_j)\, p(\mathbf{x} \mid C_j) \tag{5}$$

Applying Baye's theorem to (3), we can get equation (4) that is a condition for the decision boundary using conditional and prior probabilities. And (5) means the

discriminant function for Bayesian decision rule, if D is more than zero, a sample is classified into class C_i, else the sample belongs to class C_j.

The parameters μ, σ and ρ for skin and non-skin class are estimated using MLE (Maximum Likelihood Estimation) for about 5000 blocks (8×8 pixels) that are extracted from video clip.

3.3 Image Segmentation Using Projection

To eliminate noise and isolated lines that is made of 1 block, cross median filter applied to the binary image. After that, binary image is segmented by horizontal and vertical projection. After first projection, we get segmented regions. However, some of their regions don't contain any skin block and the region is such rough that search processing is very complex. So, further projection is needed on those regions to remove non-skin blocks, efficiently detect location and size of the face region precisely.

4 SGLD (Spatial Gray Level Dependence) Matrix and CART (Classification and Regression Tree)

The candidates of the face region have also a part of arm and neck because they have skin color. They don't have to be included in the detected face region. So, face detection algorithm is required to locate the region with only face and to minimize false alarms. Ying Dai proposed the method that it detects face region by setting threshold value for the feature value extracted from SGLD matrix [6]. And he extracted SGLD matrix that is adequate to small face size such as 16×20 pixels. We extract SGLD matrix that is adequate to the face region size for video indexing.

4.1 Feature Value Extraction from SGLD

SGLD matrix is used in textural feature analysis. For an image with level [0, L-1], let I(i, j) be the gray level value at pixel (i, j). As shown in equation (6), the number of occurrence in two neighboring pixels displaced by a vector (m, n) for m = 1, 2, … , M and n = 1, 2, … , N can be calculated. It is denoted as $P_{ab}(m, n)$ and called as SGLD matrix. In equation (6), # means the counting number of set {I(i, j) = a, I(i+m, j+n) = b}. W and H denote the width and height of the image, respectively. SGLD matrix is normalized approximately by equation (7). Textural feature values are derived from normalized matrix $N_{ab}(m, n)$.

$$P_{ab}(m,n) = \#\{(i,j),(i+m,j+n)) \in (W \times H), \; I(i,j) = a, I(i+m,j+n) = b\} \tag{6}$$

$$N_{ab}(m,n) = P_{ab}(m,n)/(W \times H) \tag{7}$$

Textural features derived from SGLD matrix include energy, entropy, inertia, inverse difference, and correlation. We utilize inertial, inverse difference, correlation shown in equation (8), (9), and (10), respectively.

$$B_I(m,n) = \sum_{a=0}^{L-1} \sum_{b=0}^{L-1} (a-b)^2 N_{ab}(m,n) \tag{8}$$

$$B_D(m,n) = \sum_{a=0}^{L-1} \sum_{b=0}^{L-1} \frac{1}{1+(a-b)^2} N_{ab}(m,n) \tag{9}$$

$$B_C(m,n) = \frac{1}{\sigma^2} \sum_{a=0}^{L-1} \sum_{b=0}^{L-1} (a-\mu)(b-\mu) N_{ab}(m,n) \tag{10}$$

To maximize classification performance using the textural features of SGLD matrix, we experiment classification into face and non-face image, varying normalized image size and the values of M and N in SGLD matrix. We evaluated the classification performance as the recall and precision ratio using cross-validation method, when learning samples are classified by classification tree. 400 face images from AT&T face database are used as samples of the face class and about 300 images for non-face class are extracted from Corel images.

Classification performance, in our experiment, varies depending on M,N, and normalized image size. The more normalized image size and the values of M and N are increased, the more classification performance is improved. However, when image size is 92×112, the performance start going worse. On the other hand, normalized image size has a little effect on SGLD extraction speed. So, we select normalized image size with maximum performance and set the value of M, N that satisfy the performance as well as the speed. When SGLD matrix with M=N=6 is extracted from 60×78 normalized image, the result of classification has the highest performance.

4.2 Classification Trees

Tree classifier is constructed to detect precise face location and size in the segmented region. To construct classification tree, we utilize $B_I(m, n)$, $B_D(m, n)$, and $B_C(m, n)$ arrays as the feature values. Classification tree has fast learning and classification procedure because of the non-metric method unlike Bayesian classifier [7]. Using binary classification tree of CART (Classification And Regression Tree) that is proposed by Breiman, Friedman, et al., classifier is learned to minimize the false detected face region.

Classification tree consists of the root node, links or branches, and terminal or leaf nodes. It follows tree-growing methodology known as CART and is grown through the splitting and pruning process [8]. In splitting process, each node chooses a feature decreasing the impurity as much as possible in the descendent node and splits the data into subsets by the selected feature. We measure impurity using Gini impurity function shown as equation (11) that is generalized as the variance impurity function useful in the two-class case. $P(\omega_i)$ of (11) is the fraction of patterns at node N that are in class ω_i.

$$i(N) = \sum_{i \neq j} P(\omega_i) P(\omega_j) = 1 - \sum_j P^2(\omega_j) \tag{11}$$

The tree growing goes on until all terminal nodes are pure. However, when the new samples are inputted to the classification tree, the performance of classifier is not

stable because it is learned by only observed samples. To settle this problem, CART methodology has the pruning procedure lowering performance for high stability to new sample.

5 PCA (Principal Component Analysis)

Feature extraction is an important step toward face identification. In previous techniques, the systems measured nodal points on the face, such as the distance between the eyes, the shape of the cheekbones and other distinguishable features. These nodal points are then compared to the nodal points computed from a database of pictures in order to find a match. But, these systems were limited based on the angle of the face captured and the lighting conditions present. New technologies are using statistical properties of the features of the images.

If a face image I(x, y) be a two-dimensional N by N array of intensity values, it can be represented by a vector of dimension N^2. Since, when normalized image size is 60×78, the classification performance is maximum, we use this face size for PCA. So, a typical image of size 60 by 78 describes a vector of dimension 4,680, or, equivalently, a point in 4,680-dimensional space. In such case, an image, maps to a point in this huge space. Images of faces will not be randomly distributed in this huge image space, rather can be represented by a relatively low dimensional subspace. The main idea of getting low dimensional subspace is to extract the important features from the face space.

This can be done using the covariance properties of the dataset. We use principal components analysis (PCA) also called the Karhunen-Loève transform for achieving low dimensional subspace. PCA is a linear transformation that chooses a new coordinate system for the data set for simplification providing the greatest variance on the first axis (called the first principal component), the second greatest variance on the second axis, and so on. PCA works in the following way in our face identification approach.

Let the training set of face images be the vectors Γ_1, Γ_2, Γ_3 ... Γ_M. The average face of the set is defined by $\Psi = \frac{1}{M}\sum_{i}^{M}\Gamma_i$. The covariance matrix is calculated as $C = \frac{1}{M}\sum_{n=1}^{M}\Phi_n\Phi_n^T = AA^T$ where $\Phi_i = \Gamma_i - \Psi$ and $A = [\Phi_1\Phi_2\cdots\Phi_M]$. Let, u_i and λ_i be the eigenvectors and associated eigen values of the C matrix respectively. But, AA^T is impractical. So, if we consider v_i as eigen vectors of the A^TA, then we can find u_i=Av_i. Thus M largest orthonormal vectors u_i and their associated values λ_i are taken for processing because they can describe the images significantly. Thus the calculations are greatly reduced from the order of the number of pixels in the images (N^2) to the order of the number of images in the training set (M). In face recognition arena, u_i is called eigenfaces. Each face image can be represented by combining these vectors linearly by equation (12) and normalized training face by equation (13)

$$\hat{\Phi} = \sum_{j=1}^{K} w_j u_j, \qquad (where \quad w_j = u_j^T\Phi) \tag{12}$$

$$\Phi_i = \begin{bmatrix} w_1^i & w_2^i & \cdots & w_K^i \end{bmatrix}^T, \qquad i = 1,2,\cdots,M \tag{13}$$

These w_j's are the extracted features for face recognition.

6 MPM (Minimax Probability Machine)

When constructing a classifier, the probability of correct classification of future data points should be maximized. There are many classifiers which try to maximize the correct classification. In face recognition, two kinds of well-known methods are Neural Network and Support Vector Machine. As a famous machine learning technique, NN has been widely used in face recognition investigated and its application on face recognition can be found in [10]. In recent years SVM, a currently popular classification technique, has been shown with many successful applications in the pattern recognition field [11]. SVM attempts to find the optimal decision boundary which separates the data points with a maximum margin based on the Structural Risk Minimization principle [11]. The research work in [12] demonstrated the success of eigenface, SVM, NN on face recognition applications. On the other hand, MPM is a very new classification technique proposed in [13, 14]. It enjoys competitive classification performance comparing with most of state-of-the-art classification techniques. The most attractive properties of MPM are that it can explicitly provide a worst-case bound on the probability of misclassification of future data when the mean and covariance matrix of the data are known. The basic theory of MPM for binary classification is discussed below.

The goal of MPMC is to find a decision boundary $H(a,b)=\{z|\mathbf{a}^T z = b\}$ such that the minimum probability Ω_H of classifying future data correctly is maximized. If we assume that the two classes are generated from random vectors $\mathbf{x}$ and $\mathbf{y}$, we can express this probability bound just in terms of the means and covariance of these random vectors. The following result due to Marshall and Olkin [15] and extended by Bertimas and Popescu [16] provides the theoretical basis for assigning probability bounds to hyperplane classifiers:

$$\Omega_H = \inf_{\mathbf{x}\sim(\bar{x},\Sigma_\mathbf{x}),\mathbf{y}\sim(\bar{y},\Sigma_\mathbf{y})} \Pr\{\mathbf{a}^T\mathbf{x} \geq b \wedge \mathbf{a}^T\mathbf{y} \leq b\} \tag{14}$$

Note that we do not make any distributional assumptions other than that $\bar{\mathbf{x}}, \Sigma_\mathbf{x}$, $\bar{\mathbf{y}}$, $\Sigma_\mathbf{y}$ are bounded. Exploiting a theorem from Marshall and Olkin[15], it is possible to rewrite(14) as a closed expression: $\Omega_H = \frac{1}{1+m^2}$ where

$$m = \min_{\mathbf{a}} \sqrt{\mathbf{a}^T\Sigma_\mathbf{x}\mathbf{a}} + \sqrt{\mathbf{a}^T\Sigma_\mathbf{y}\mathbf{a}} \qquad s.t. \qquad \mathbf{a}^T(\bar{\mathbf{x}} - \bar{\mathbf{y}}) = 1 \tag{15}$$

The optimal hyper plane parameter $\mathbf{a}^*$ is the vector that minimizes (15). The hyper plane parameter b* can then be computed as equation (16).

$$\mathbf{b}_* = \mathbf{a}_*^T \overline{\mathbf{x}} - \frac{\sqrt{\mathbf{a}_*^T \Sigma_{\mathbf{x}} \mathbf{a}_*}}{m} \tag{16}$$

A new data point Z_{new} is classified according to sign($\mathbf{a}_*^T Z_{new} - \mathbf{b}_*$); if this yields +1, Z_{new} is classified as belonging to class x, otherwise it is classified as class y.

7 Experiments

In this section we report the results of experiments that we carried out to test our algorithmic approach. We experiment three different videos for face region detection. Our system shows precision ratio of 0.85% in an average for face region detection. The performance in different videos varies due to the variations of video quality. Fig. 2 shows the extracted face regions from several video frames.

We, in our experiment, build a database of 400 face images, including 40 distinct persons, each with 10 faces that vary in position, rotation, scale, and different expressions. Fig. 3 shows snapshots of two persons from constructed face database.

Fig. 2. Detected face regions

Fig. 3. Snapshots of an actress and an actor from face database

For multi-class pattern recognition, we construct multi-class classification technique by employing binary MPM classifier. The database is partitioned into 10 subsets; each contains 1 face image from each distinct person. Like face detection, here we use the 10-folder cross validation also. Ten trails are run in the experiment. Table 1 shows the final experimental results for AT&T face database and video sequence. It contains the result of identification after applying face region detection in the video frames. The result in percentage represents the correct identification rate.

Table 1. Experimental results on the AT&T face database and video sequence

Subset	AT&T Face Database	Video sequence
1	93.37 %	90.89 %
2	95.06 %	91.91 %
3	96.06 %	91.93 %
4	95.31 %	90.45 %
5	94.43 %	88.88 %
6	95.87 %	90.9 %
7	96.68 %	91.91 %
8	97.93 %	92.96 %
9	94.37 %	89.98 %
10	95.87 %	88.56 %
Average	95.30 %	90.84 %

8 Conclusion and Future Work

We propose an efficient actor dependent video retrieval method using classification tree for face region detection which is followed by Minimax Probability Machine for recognition. Classification tree was constructed by facial texture features from SGLD matrix. For efficiency of search procedure, we used the skin detection method of Bayesian decision rule and segmented image by horizontal and vertical projection. For recognition, we use PCA for feature extraction and a state-of-the-art classification technique called Minimax Probability Machine (MPM). Although, PCA and MPM have higher computational cost, it can reliably evaluate future data class. MPM uses means and covariance matrices for classification. Therefore, correct estimation of the means and covariance matrices is important for face recognition. Experimental result shows that classification rates are 95.30% and 90.84% for AT&T face database and video frames, respectively. Although, video frames suffers from variations in lighting, pose, scale etc our system shows good result. We get very good result if we use face region detection and then recognition.

In future work, we will consider event based video summarization and indexing by character recognition with the proposed method.

References

1. Azriel Rosenfeld et al., Video Mining, Kluwer Academic Publishkers, 2003.
2. Minerva Yeung et al., “Segmentation of Video by Clustering and Graph Analysis”, Computer Vision and Image Understanding, vol. 71, no. 1, pp. 94~109, 1998.
3. Hualu Wang et al. "A Highly Efficient system for Automatic Face Region Detection in MPEG Video", IEEE Transactions on Circuit and Systems for Video Technology, vol. 7, no 4, pp. 615-628, 1997.

4. Ying Li et al., Video Content Analysis using Multimodal Information, Kluwer Academic Publishers, 2003.
5. Ming-Hsuan Yang et al., "Detecting Faces in Images: A Survey", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 1, pp. 34-58, 2002.
6. Ying Dai and Yasuaki Nakano, "Face-Texture Model Based On SGLD And Its Application in Face Detection in a Color Scene", Pattern Recognition, vol. 29, no. 6, pp. 1007-1017, 1996.
7. Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification, Wiley-Interscience, 2001.
8. Breiman and Friedman et al., Classification And Regression Trees, CRC Press, 1998.
9. http://www.uk.research.att.com
10. J. Zhang, Y. Yong, and M Lades, “Face recognition: eigenface, elastic matching, and neural nets,” Prof. of the IEEE, vol. 85, pp. 1423-1435, 1997.
11. V.N. Vapnik, Statistical Learning Theory, Wiley, 1998.
12. Phillips, “Support vector machines applied to face recognition,” in Proc. Advances in Neural Information Processing Systems, 1998.
13. G.R.G. Lankriet, L. El Ghaoui, C. Bhattacharyya, and M. I. Jordan, “Minimax Probability Machine,” in Proc. Advances in Neural Information Processing Systems, 2002.
14. G.R.G. Lankriet, L. El Ghaoui, C. Bhattacharyya, and M. I. Jordan, “A robust minimax approach to classification,” Journal of Machine Learning Research, vol. 3, pp. 552-582, 2002.
15. W. Marshall and I. Olkin. Multivariate chyshev inequalities. Annuals of Mathematcal statistics, 31(4)L1001-1014, 1960.
16. Popescu and D. Bertsimas. Optimal inequalities in probability theory: A convex optimization approach. Technical Report TM62, INSEAD, Dept. math. O.R., Cambridge, Mass, 2001.
17. G.R.G. Lanckriet, L. El Ghaoui, C. Bhattacharyya, and M. I. Jordan, “A robust minimax approach to classification”, Journal of Machine Learning Research, vol. 3, pp. 552-582, 2002.

ART-Artificial Immune Network and Application in Fault Diagnosis of the Reciprocating Compressor

Maolin Li, Na Wang, Haifeng Du, Jian Zhuang, and Sun'an Wang

Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System
Xi'an Jiaotong University, Xi'an 710049, China
maolinli@mail.xjtu.edu.cn

Abstract. Inspired by complementary strategies, a new fault diagnostic method, which integrates with the Adaptive Resonance Theory (ART) and Artificial Immune Network (AIN), is proposed in this paper. With the help of clustering of ART neural network, the vaccines that image features of data set are extracted effectively, and then an AIN named aiNet is adopted to realize data compression. Finally the memory antibodies optimized by aiNet can be used to recognize each feature of original dataset and to realize fault diagnosis. The experimental results show that the approach is useful and efficient for the fault diagnosis of the multilevel reciprocating compressor.

1 Introduction

The multilevel reciprocating compressor is a system with multifactor, coupling and nonlinear. So it is difficult to establish the precise mathematic model. In addition, due to its fault information is numerous and complicated, many conventional methods maybe can find the faults, but cannot give the reasons [1]. To some extent, these methods are limited in the application. However, some new methods have been put forward to solve these similar problems based on artificial intelligence strategies.

As an artificial intelligence method, Adaptive Resonance Theory (ART) neural network has some good properties, such as online learning and nonstationary signal process, etc. Because of recognizing the samples steadily and quickly, ART is suitable for clustering and classification [2]. In addition, artificial immune system (AIS) describes a learning technique based on the heuristic of biological immune system and the natural defense mechanism of learning outside material. Due to integrating with many mechanisms of noise tolerance, un-supervise learning, self-organizing and immune memory, etc., AIS has the potential to solve many complex problems [3][4].

In this paper, inspired by complementary strategies, a novel ART-Artificial Immune Network model, namely ART-aiNet algorithm, is proposed. With the excellent clustering ability of ART, the prior information of patterns can be obtained, such as the classification and data distributing. Then, in order to accelerate the evolution of antibody, the centers of clustering are taken as vaccines that inoculate to be initial antibodies. Thereafter, under the optimization of aiNet, not only the computation effort decreases, but also the features of dataset can be obtained. Finally, the optimization memory antibodies can distinguish the different data patterns.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 502 – 505, 2006.

The rest of this paper is organized as follows. In Section 2, ART-aiNet algorithm is introduced. Section 3 presents the fault diagnosis experiments of the four-level reciprocating compressor and gives the conclusions.

2 ART-Artificial Immune Network

For convenience, the proposed method adopts definition of paper [5]. $\boldsymbol{Ag}$ is the population of antigens, and $\boldsymbol{Ag}_i$ is the input antigenic pattern with N_p. $\boldsymbol{Ab}$ is the population of N_t network cells (antibodies) with elements $\boldsymbol{Ab}_j$, and $\boldsymbol{Ab}_{\{n\}}$ is the total memory antibody repertoire. $\boldsymbol{C}$ is the population of N_c clones generated from $\boldsymbol{Ab}$, and $\boldsymbol{C}^*$ is the population $\boldsymbol{C}$ after the affinity maturation process. The memory antibodies set $\boldsymbol{M} \in \boldsymbol{Ab}$, has n memory antibodies with p-dimension. $\boldsymbol{D}$ is the affinity matrix between each pair $\boldsymbol{Ag}_i$-$\boldsymbol{Ab}_j$ with elements d_{ij} (i = 1,…, N_p, j = 1,…, N_t) which reflects the probability of starting a clonal response. Similarly, the affinity matrix $\boldsymbol{S}$ will be assigned to each pair $\boldsymbol{Ab}_i$-$\boldsymbol{Ab}_j$ (i, j = 1,…, N_t), with elements s_{ij}, reflecting their similarity. ξ is the percentage of the mature antibodies to be selected, and $\sigma_{d,s}$ is the natural death and suppression threshold, respectively. Meanwhile, the suppression threshold σ_s, which performed by eliminating the self-recognizing antibodies, is determined by the affinity of vaccines, and the natural death threshold σ_d is given by experience.

The ART-aiNet learning algorithm can be described as follows:

Step 1 Data preprocess: there are filter process and outlier removing;
Step 2 The ART is adopted to cluster the patterns, and the centers of classes are used as initial antibodies. Meanwhile, σ_s is determined by the minimal affinity s_{ij};
Step 3 Test stopping criterion:
Step 3.1 For each antigenic pattern $\boldsymbol{Ag}_i$, do:
Step 3.1.1 Determine its affinity d_{ij} to all $\boldsymbol{Ab}_j$, j = 1,…, N_t;
Step 3.1.2 A subset $\boldsymbol{Ab}_{\{n\}}$ composed of the n highest affinity antibodies is selected;
Step 3.1.3 The n selected antibodies are going to proliferate (clone) proportionally to their antigenic affinity d_{ij}, generating a set $\boldsymbol{C}$ of clones;
Step 3.1.4 The set $\boldsymbol{C}$ is submitted to a directed affinity maturation process (guided mutation) generating a mutated set $\boldsymbol{C}^*$.
Step 3.1.5 Determine the affinity $\boldsymbol{D}$ among $\boldsymbol{Ag}$ and all the elements of $\boldsymbol{C}^*$;
Step 3.1.6 From $\boldsymbol{C}^*$, re-select ξ % of the antibodies with highest affinity and put them into a matrix $\boldsymbol{M}$ of clonal memory;
Step 3.1.7 Eliminate all the memory clones from $\boldsymbol{M}$ whose affinity less than σ_d;
Step 3.1.8 Determine the affinity s_{ij} among the memory clones;
Step 3.1.9 Clonal suppression: eliminate those memory clones whose $s_{ij} < \sigma_s$;
Step 3.1.10 Concatenate the total antibody memory matrix with the resultant clonal memory $\boldsymbol{M}^*$ for $\boldsymbol{Ag}_i$: $\boldsymbol{Ab}_{\{n\}} \leftarrow [\boldsymbol{Ab}_{\{n\}}; \boldsymbol{M}^*]$;
Step 3.2 Network suppression: eliminate all the antibodies such that $s_{ij} < \sigma_s$;
Step 3.3 Eliminate a percentage d % of randomly generated antibodies;
Step 4 Return Step 3.

In the above algorithm, the affinity between two immune antibodies is represented by Euclidean distance in Step 3.1.1, 3.1.5 and 3.1.8. Specially, the ***Ag-Ab*** affinity is inversely proportional to the distance between them: the smaller the distance, the higher the affinity, and vice versa. In addition, in Step 3.1.4, in order to determine the mutation process of antibodies, the following equation was employed

$$\boldsymbol{C}^{*} = \boldsymbol{C} - \alpha f(\boldsymbol{C} - \boldsymbol{Ag}) . \tag{1}$$

Where $\boldsymbol{C}$ is the antibody matrix, $\boldsymbol{Ag}$ is the antigenic matrix. The parameter α is the learning rate which inversely proportional to the antigenic affinity of its parent antibodies. In addition, the transcendental function $f(x)$ is given by

$$f(x) = \mathrm{sgn}(x)\exp(x) . \tag{2}$$

The network outputs can be taken to be the matrix of memory antibodies ($\boldsymbol{M}$) and their matrix of affinity ($\boldsymbol{S}$). While the matrix $\boldsymbol{M}$ represents the network internal images of the antigens presented to the aiNet, matrix $\boldsymbol{S}$ is responsible for determining which network antibodies are connected to each other, describing the general network structure. Therefore, the memory antibodies matrix $\boldsymbol{M}$ reflects the features of input patterns, and can be used directly for pattern recognition, or can used as input information for other diagnostic strategies.

3 Experiments and Conclusions

The fault diagnosis of the multilevel reciprocating compressor is still an unsolved problem. The valve, which is a vulnerable part in a compressor, can directly influence the air displacement, power consumption and the safety of the compressor. The outlet valve leakage may result in temperature rise in output air and change in air pressure, these make the compressor cannot work in normal.

For a four-level compressor shown in Fig. 1, the samples are collected respectively in the three situations: outlet valve failure in the first level, inspiratory valve failure in the second level and normal. Thereafter, 3,586 samples are obtained by some preprocess methods (filter, gross errors elimination and normalization). Each sample contains eight variables, which are pressure variables p_1, p_2, p_3 and p_4 and temperature variables T_1, T_2, T_3 and T_4.

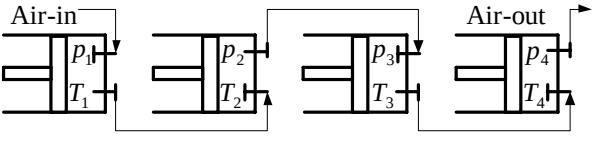


Fig. 1. Sketch of a four-level reciprocating compressor

The contrastive results obtained by ART-aiNet approach and Rough Set (RS) method [6] are shown in Table 1. Meanwhile, an index E_n is adopted to evaluate the influence of accuracy of decision rules concluded by the data compression [7].

Table 1. The contrastive results

Method	ART-aiNet	RS
Classification success rate	95.5%	88%
Decision rules	195	88
Evaporation ratio E_n	90.28%	83.68%

From Table 1, it can be seen, that the number of decision rules obtained by ART-aiNet is more than Rough Set-based method, but the classification success rate of dataset and the evaporation ratio E_n based on ART-aiNet are better. So the ART-aiNet approach shows superior performance in diagnosis classification success rate and data evaporation ratio.

In addition, compared with other immune algorithms, the ART-aiNet can decrease the computational cost, but it still cannot satisfy the online monitoring. Thereby, there is need to search for some new strategies to decrease computational cost in further.

Acknowledgements

The support of the National Natural Science Foundation of China (NSFC) (No. 50505034) is gratefully acknowledged.

References

1. Ren, Q. M., Ma, X. J., Miao, G.: Application of Support Vector Machines in Reciprocating Compressor Valve Fault Diagnosis. Lecture Notes in Computer Science, Vol. 3611. Springer-Verlag, Berlin Heidelberg New York (1997) 81–84
2. Dunbar, G.: The Clustering of Natural Terms: an Adaptive Resonance Theory Model. IEEE International Joint Conference on Neural Network. 6 (1999) 4362–4364
3. Timmis, J., Neal, M., Hunt, J.: An Artificial Immune System for Data Analysis. Biosystems. 55 (2000) 143–150
4. Luh, G. C., Cheng, W. C.: Immune Model-Based Fault Diagnosis. Mathematics and Computers in Simulation. 67 (2005) 515–539
5. Castro, L. N., Zuben, F. J.: aiNet: An Artificial Immune Network for Data Analysis. Idea Group Publishing, Pennsylvania (2001)
6. Wang, J., Wang, R., Miao, D.Q., et al.: Data Enriching Based on Rough Set Theory. Chinese Journal of Computers. 21 (1998) 393–400
7. Du, H.F., Wang, S.A.: Fault Diagnose of the Reciprocating Compressor Based on ART-Artificial Immune Network. Chinese Journal of Mechanical Engineering. 38 (2002) 88–90

Online Composite Sketchy Shape Recognition Based on Bayesian Networks

Zhengxing Sun, Lisha Zhang, and Bin Zhang

State Key Lab for Novel Software Technology, Nanjing University, PR China, 210093
szx@nju.edu.cn

Abstract. This paper presents a novel approach for online multi-strokes composite sketchy shape recognition based on Bayesian Networks. By means of the definition of a double-level Bayesian networks, a classifier is designed to model the intrinsic temporal orders among the strokes effectively, where a sketchy shape is modeled with the relationships not only between a stroke and its neighbouring strokes, but also between a stroke and all of its subsequence.. The drawing-style tree is then adopted to capture the users' accustomed drawing styles and simplify the training and recognition of Bayesian network classifier. The experiments prove both effectiveness and efficiency of the proposed method.

1 Introduction

As sketching can be to help us convey ideas and guide our thinking process both by aiding short-term memory and by helping to make abstract problems more concrete in the graphic computing [1], numerous researchers have been working on the subject of online sketch recognition for many years. It is true that some prototype systems come into being in online sketch recognition either as a natural input modality [2][3][4] or to recognize the composite sketchy shapes [5][6]. However, sketching is usually informal, inconsistent and ambiguous. Most of the existing methods are either only applicable to single-stroke sketches (such as rectangles and circles) [4] or limited to their computational complexity and high sensitivity to the segmentation process [5][6]. More important, this brings on the poor effect and efficiency of sketch recognition engines, especially for multi-strokes sketchy shapes and newly added users. Therefore, adaptive sketchy shape recognition is anticipant, where recognition engine should be trainable and adaptable to a particular user's drawing styles, and have a priori probabilities of the sketchy shape classes for each user [7].

Obviously, one solution for adaptive sketch recognition is to construct appropriate classifiers based on machine learning to learn users' drawing styles. As strokes are natural conceptual elements of freehand drawings, modeling sketchy shapes with strokes is important for classifiers. The conventional approaches convert a drawing input into local feature vectors, assume independences between them, and use statistical or structural classifiers such as Support Vector Machine (SVM) and Hidden Markov models (HMM) in classifying them. In our previous researches, we have developed a stroke classification method based on SVM [7][8], where a modified

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 506 – 515, 2006.

turning function is used to model sketchy shapes, and virtual strokes are introduces to link continuous strokes orderly and translate the multi-strokes sketchy shapes into a single stroke. Recently, HMM has been used for online sketch recognition [8][10], where the drawing patterns are treated as the result of a stochastic process that is governed by a hidden stochastic model and identified according to its probability of generating the output, inspired by its success in speech recognition and handwriting recognition. However, the dimension of feature vectors in SVM classifier must be fixed for all shapes and the number of states in HMM classifier depends heavily on stabilization of drawing styles. Moreover, relationships between strokes cannot felicitously be modeled (In fact, stroke relationships are indispensable to discriminate between sketches of similar pen-movements, and only adjacent relationships between temporal-sequence strokes are considered). These limit their capability to capture users' drawing styles.

In this paper, we propose to explicitly model drawing sketches with strokes and their relationships in a probabilistic framework. Bayesian networks, a well-known framework for modeling dependencies, are adopted to represent drawing sketch models whose nodes correspond to strokes and whose arcs their dependencies. A decision tree is then defined to model and capture users' drawing styles of special symbols or shapes. The rest of the paper is organized as follows. In section 2, the framework and its principle of our Bayesian Networks framework for adaptive online sketch recognition is introduced in detail. In section 3, some experimental results are evaluated and conclusions are given in the final section.

2 Bayesian Networks Framework for Sketch Recognition

A Bayesian network [11] is a graph with probabilities for representing random variables and their dependencies. It efficiently encodes the joint probability distribution of a large set of variables. Its nodes represent random variables and its arcs represent dependencies between random variables with conditional probabilities at nodes. It is a Directed-Acyclic Graph (DAG) so that all edges are directed and there is no cycle when edge directions are followed. Using a Bayesian Network offers many advantages over traditional methods of determining causal relationships [11]. Independence among variables is easy to recognize and isolate while conditional relationships are clearly delimited by a directed graph edge: two variables are independent if all the paths between them are blocked. Not all the joint probabilities need to be calculated to make a decision; extraneous branches and relationships can be ignored. By optimizing the graph, every node can be shown to have at most k parents. The algorithmic routines required can then be run in $O(2^k n)$ instead of $O(2^{n)}$ time. In essence, the algorithm can run in linear time (based on the number of edges) instead of exponential time (based on the number of parameters).

2.1 Bayesian Networks Structure for Sketch Recognition

There have been few researches using Bayesian network for online graphics recognition. Cho and Kim [12] have proposed a Bayesian network framework for explicitly modeling components and their relationships of Korean Hangul characters. A character is modeled with hierarchical components: a syllable model, grapheme models, stroke

models and point models. However, it is not enough for adaptive sketch recognition to model drawing styles with a point model represented by a 2-D Gaussian for point positions on X-Y plane. Alvarado et al [13] have described a model for dynamically constructing Bayesian network to represent varying hypotheses for the task of sketch recognition and allow both stroke data and contextual data to influence the probability of an interpretation for the users' strokes. However, they put their emphasis upon drawing semantic description and inference for sketch understanding to assist sketch recognition, not for adaptive sketch recognition (Their sketch recognition process is not actually brought into Bayesian network framework). For adaptive online sketch recognition, we design a Bayesian Network structure as shown in **Fig. 1**.

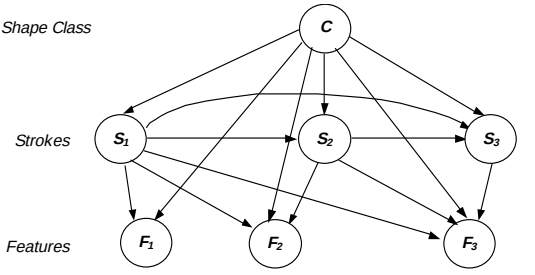


Fig. 1. Bayesian Networks Structure for Adaptive Sketch Recognition

We consider three types of random variables: two discrete variables - the shape classes ***C*** and stroke types ***S***, one continue variable - feature vectors ***F***. Accordingly, all nodes in our Bayesian Network structure are arranged in three level: only one node in top level is the shape class, nodes in middle level are several strokes constituted a sketchy shape and nodes in the bottom level are some multi-dimension feature vectors correspond to each of strokes. In **Fig. 1**, node ***C*** indicates a type of sketchy shape, nodes S_1, S_2, S_3 represent respectively the first, second and third stroke drawn by user for a sketchy shape ***C***, and nodes F_1, F_2, F_3 are feature vectors respectively corresponding to the first, second and third stroke. In order to simplify Bayesian Network structure, the directed arcs in the structure link each of nodes in keeping with three regulations relative to the speciality of freehand drawing: (1). Each of strokes is only relative to a sketchy shape and all previous inputting strokes; (2). Each of feature vectors for every stroke is only relative to a sketchy shape and all previous inputting strokes; (3). There are no relationships between the feature vectors corresponding to every stroke. As shown in **Fig. 1**, there are directed arcs between node ***C*** and other nodes because all of them are composed of a sketchy shape. According to regulations, there are directed arcs between each pair of nodes $S_1 \sim S_2, S_1 \sim S_3, S_2 \sim S_3, S_1 \sim F_1$, $S_1 \sim F_2$, $S_2 \sim F_2$, $S_1 \sim F_3$, $S_2 \sim F_3$, $S_3 \sim F_3$ respectively, and no arcs between nodes F_1, F_2, F_3.

2.2 Maximum Posteriori Probability Estimation

According to our definition of Bayesian Network, adaptive sketch recognition can be seen as the problem of solution of maximum posteriori probability.

Supports there are ***m*** classes of sketchy shapes to be recognized and a sketchy shape is composed of ***n*** number of strokes, the posteriori probability can be calculated in terms of Bayes expressions as follow:

$$P(C|S_1,S_2,\cdots,S_n,F_1,F_2,\cdots F_n)=\frac{P(C,S_1,S_2,\cdots,S_n,F_1,F_2,\cdots F_n)}{P(S_1,S_2,\cdots,S_n,F_1,F_2,\cdots F_n)} \tag{1}$$

Because $P(S_1,S_2,\cdots,S_n,F_1,F_2,\cdots F_n)$ is independent on the shape classes $\boldsymbol{C}$, the problem of maximum posteriori probability means to maximize the joint probability $P(C,S_1,S_2,\cdots,S_n,F_1,F_2,\cdots F_n)$. The sketch recognition process can then be described as to find a solution to maximum joint probability as follow:

$$C_R=\arg\max P(C_j,S_1,S_2,\cdots,S_n,F_1,F_2,\cdots,F_n),\ \ j=1,2,\cdots,m. \tag{2}$$

In a Bayesian network, the joint probability of a set of random variables $\{x_1,x_2,\cdots,x_n\}$ is calculated by the multiplication of local conditional probabilities of all the nodes. Let a node x_i denotes the random variable x_i, and $par(x_i)$ denote the parent nodes of x_i, from which dependency arcs come to the node x_i. Then, the joint probability of $\{x_1,x_2,\cdots,x_n\}$ is given as follows:

$$P(x_1,x_2,...,x_n)=\prod_{i=1}^{n}P(x_i\mid par(x_i)) \tag{3}$$

Therefore, we can gain:

$$\begin{aligned}
&P(C_j,S_1,S_2,...,S_n,F_1,F_2,...,F_n)\\
&=P(C_j)*\prod_{i=1}^{n}P(S_i\mid par(S_i))*\prod_{i=1}^{n}P(F_i\mid par(F_i))\\
&=P(C_j)*\prod_{i=1}^{n}P(S_i\mid C_j,S_1,S_2,...S_{i-1})*\prod_{i=1}^{n}P(F_i|\,C_j,S_1,S_2,...,S_{i-1},S_i)\\
&=P(C_j)*\prod_{i=1}^{n}\frac{P(C_j,S_1,S_2,...,S_{i-1},S_i)}{P(C_j,S_1,S_2,...,S_{i-1})}*\prod_{i=1}^{n}P(F_i\mid C_j,S_1,S_2,...,S_{i-1},S_i)
\end{aligned} \tag{4}$$

Where, $P(F_i|\,C_j,S_1,S_2,...,S_{i-1},S_i)$ is the conditional probability between continuous random variables with high order dependencies and $P(S_i\mid C_j,S_1,S_2,...S_{i-1})$ is the conditional probability between discrete random variables.

We suppose that the feature vectors of a drawing satisfy the random variables Gaussian distributions, that is to say, $P(F_i\mid C_j,S_1,S_2,...,S_{i-1},S_i)\sim N(\mu,\Sigma)$. Therefore,

the conditional probability $P(F_i|C_j, S_1, S_2,..., S_{i-1}, S_i)$ can be modeled by conditional Gaussian distributions using following formula:

$$p(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)] \quad (5)$$

where, $\boldsymbol{x}$ is the feature vectors of random variables, $\boldsymbol{d}$ is the dimension of feature vectors, $|\boldsymbol{\Sigma}|$ is the determinant value of covariance matrix $\boldsymbol{\Sigma}$ of feature vectors $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma^{-1}}$ is the converse matrix of covariance matrix $\boldsymbol{\Sigma}$.

The conditional probability $P(S_i|C_j, S_1, S_2,...S_{i-1})$ can be estimated by the frequency distribution in sample space as follow:

$$P(S_i|C_j, S_1, S_2,..., S_{i-1}) = \frac{P(C_j, S_1, S_2,..., S_i)}{P(C_j, S_1, S_2,..., S_{i-1})} = \frac{Count(C_j, S_1, S_2,..., S_i)}{Count(C_j, S_1, S_2,..., S_{i-1})} \quad (6)$$

where, $Count(C_j, S_1, S_2,..., S_{i-1})$ is the total number of samples which is composed of a series of strokes $S_1, S_2,..., S_{i-1}$ and belongs to the shape class C_j, $Count(C_j, S_1, S_2,..., S_{i-1}, S_i)$ is the total number of samples which is composed of a series of strokes $S_1, S_2,..., S_{i-1}, S_i$.

The priori probability $P(C_j)$ can be estimated simply by the frequency distribution in sample space as follow

$$P(C_j) = Count(C_j)/Count(AllSamples) \quad (7)$$

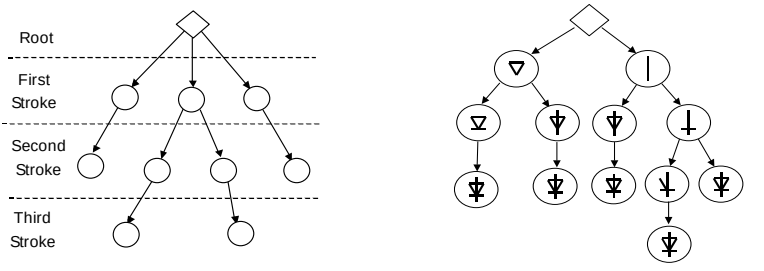
Where, $Count(C_j)$ is the total number of samples belonged to the shape class C_j and $Count(AllSamples)$ is the total number of samples.

2.3 Drawing Style Modeling

Theoretically, the conditional probabilities are computable. However, they require a high computing complexity and a large of posteriori probability matrix, for example, $P(S_i|C_j, S_1, S_2,...S_{i-1})$ requires reasoning M^n cases and $(N_1 \times N_2 \times \cdots \times N_n)$-dimensions matrix for $\boldsymbol{M}$ classes of shape with $\boldsymbol{n}$ number of strokes.

In fact, the posteriori probability matrix is sparse for a specific user, because his/her drawing styles (including the number and type of strokes, the temporal order of primitives and so on) for a shape is relatively fixed though the possible drawing styles of a shape may be diverse [7]. Therefore, we define a particular data structure to model drawing styles (the probability value in the posteriori probability matrix for sketch recognition is nonzero) of a shape for a specific user, as shown in Fig. 2(a), named as drawing-style tree (DS-tree) to be used as the model of Bayesian classifier training and recognition. A drawing-style tree is hierarchical according to the drawing sequences of strokes and enumerates all accustomed drawing styles/manner of a shape for a specific user, where the root represents a class of shape, every children node indicates the addition of a stroke and the attributes of drawing till current stroke,

every directed arc represents the successive relationships between two strokes, each of level lists all possible type of strokes in a drawing step and each of branches represents a possible drawing style/sequence consisted of a series of strokes. Fig. 2(b) shows an example of a user accustomed drawing styles/manner of electronic diode symbol in term of drawing-style tree for a specific user.



(a) Definition of drawing-style tree (b) A drawing style tree of diode symbol

Fig. 2. Illustration of Drawing-Style Tree

The attributes of drawing in each node include the type of current stroke and its corresponding feature vectors, the list of previous strokes and their corresponding feature vectors, the statistical access frequency of drawing style along with a branch from root to current node (that is, the value of $Count(C_j, S_1, S_2, ..., S_{i-1}, S_i)$), as well as the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ of feature vectors corresponding to current drawing.

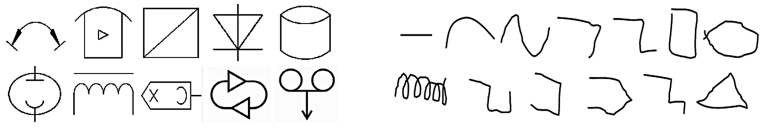
Based on definition of drawing-style tree, the task of Bayesian Networks classifier training to establish a drawing-style tree for each of shape classes by calculating and updating iteratively the value of $Count(C_j, S_1, S_2, ..., S_i)$ and the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ of the corresponding node in term of the collected samples. For every stroke of sample, if there is node in every level of the drawing-style tree so that the type of the current stroke and the list of previous strokes at this node are same as that of sample, increment the number of statistical access frequency of drawing style or the value of $Count(C_j, S_1, S_2, ..., S_i)$ by one, add the feature vectors of this sample to node and update the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ of feature vectors of drawing in this node. If not, create a new node in current level of the drawing-style tree, calculate the attributes of drawing in this novel node according to the stroke of sample and set the statistical access frequency of drawing style as one.

After established of drawing-style tree, the adaptive sketch recognition based on Bayesian Networks classifier for each of users' online drawing can be describe as follow. For each of shape classes, the following probabilities are calculated: (a) the priori probability $P(C_j)$ by means of equation (7) for root node, (b) the conditional probability $P(F_i | C_j, S_1, S_2, ..., S_i)$ according to equation (5) and the conditional probability $P(S_i | C_j, S_1, S_2, .. S_i)$ according to equation (6) for each of nodes, (c) the joint probability $P(C_j, S_1, S_2, ..., S_n, F_1, F_2, ..., F_n)$ according to equation (4) for each of

branches in a drawing-style tree, and (d) the maximal joint probability $P(S_i)=P(C_j)\times P(C_j,S_1,S_2,...,S_n,F_1,F_2,...,F_n)$ for each of shape classes. For all shape classes, some candidate drawing styles or candidate recognition results can then be sorted by the magnitude of the maximal joint probability $P(S_i)$ of each of classes.

3 Experimental Results and Evaluation

In our experiments, we select 10 categories of typical symbols from the Chinese Electric Symbols Standard (GBT4728&GBT5465), which are most commonly used in circuit diagramming, as shown in Fig. 3(a). Some of them are simple in structure and others relatively complicated. We design two experiments to validate the effects of our proposed method in contrast with our two previous method: SVM classifier [7][8] and HMM classifier [10]. All experiments are run on Microsoft Windows XP with 1.4 GHz Intel CPU and 512MB memory).



(a) Categories of graphical symbols (b) Some types of drawing strokes

Fig. 3. Symbols and strokes used in our experiments

The goal of first experiment is to compare the convergent performances of three classifiers under different sample sets. We collect 1000 samples for each of the 10 graphical symbols, which are divided into a training set and a testing set with the proportion of 7 to 3. We do the experiment six times, with the total samples of 100, 200, 400, 600, 800 and 1000, respectively. Thus, in all these experiments, the sizes of training samples are in turn 70, 140, 280, 420, 560 and 700. All samples are represented as the modified turning function [7] with 20-dimensional features. **Fig. 4** shows the recognition precisions of three classifiers training under different sizes of sample sets in our first experiment and **Fig. 5** shows the training time of three classifiers under different sizes of sample sets in our first experiment.

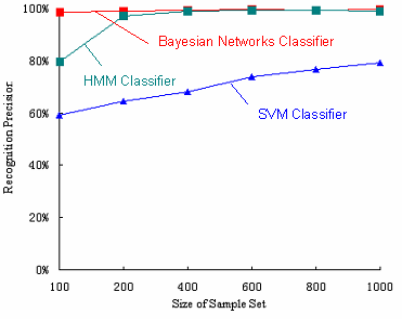
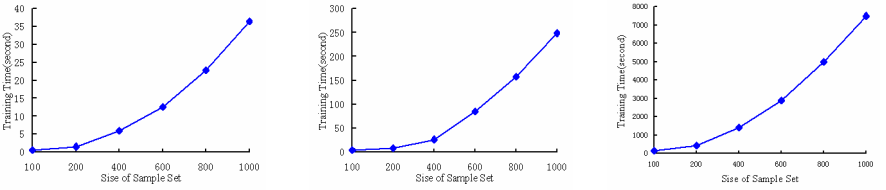


Fig. 4. Recognition precision of three classifiers under different size of samples



(a) Bayesian Classifier (b) SVM Classifier (c) HMM Classifier

Fig. 5. Training time of three classifiers under different sizes of sample sets

From **Fig. 4** we can see that the recognition precision of both HMM and Bayesian Networks classifiers reaches a convergence of perfect precision at a small sample sets (For HMM classifiers, the recognition precision reaches 97% at 400 samples, for Bayesian Networks classifier, it reaches 99% at 100 samples) while the recognition precision of SVM classifiers is poor. That is, both HMM and Bayesian Networks classifiers are more suitable for multi-stroke symbols recognition than SVM classifier. From **Fig. 5** we can see that the training time of SVM and Bayesian Networks classifiers is acceptable while that of HMM classifier is tremendous.

The second experiment is designed to verify the multi-users adaptability of three classifiers under different feature representations. We invite two users to draw 400 and two users to draw 500 samples for every graphical symbol, where three draw in only one manner they prefer themselves while one draws in several manners so that there are more than one kind of combinations of strokes in a symbol' samples by using some types of strokes as shown in **Fig. 3**(b). The samples of each user are also divided into training samples and testing samples according to the proportion of 7 to 3. All samples are represented as the modified turning function [7] and the composite feature vectors [10] respectively. **Table 1** lists the recognition precision of three classifiers for four users under two types of feature representation. From **Table 1**, we can see that the recognition precision of both HMM and Bayesian Networks classifiers is perfect for multi-users adaptation and insensitive to feature representation while the recognition precision of SVM is poor and sensitive to feature representation.

Table 1. Average Recognition Precisions of three classifiers for multi-users (%)

Classifier	Turning Function Feature				Composite Feature			
	User1	User2	User3	User4	User1	User2	User3	User4
SVM	70.27	68.99	63.17	68.25	43.30	56.08	30.30	30.44
HMM	92.53	84.20	92.83	99.00	90.07	94.73	83.42	95.75
Bayesian	99.27	96.82	97.16	99.42	99.53	97.82	99.67	100

As an experimental result, we can conclude that the Bayesian Networks classifier proposed in this paper is the most preferable for online adaptive multi-strokes sketch recognition both in recognition and training performances and multi-users adaptability among three classifiers, at least in our defined domain. SVM is mainly designed for distinguishing different categories and their decision regions need to be convex and are sensitive to the definition of feature spaces. This restriction limits the flexibility

and accuracy of the SVM classifiers only suitable for simple sketchy shape recognition or stroke classification [7][8]. In addition, the training of SVM classifier is actually an iterative process sensitive to feature definition [7][8]. We employ a first-order left-to-right chain for HMM classifier [515], supposed that each stroke is strongly relevant to the previous and the next. This is an appropriate HMM mode for the strictly serial drawing in sketch. In other hand, our HMM classifier for sketch recognition is sensitive to stroke orders. These may result the larger training time and unsatisfactory recognition, especially for complicated symbols. Our Bayesian Networks classifier models a sketch with the relationships not only between stroke and its neighbouring strokes, but also between a stroke and all of its subsequence. Therefore, it may be more suitable to describing multi-stroke symbols than SVM and HMM. As shown in **Table 1**, when the three classifiers work on user4's samples which are drawn in more than one manner, the recognition result of Bayesian Networks classifier is much more prominent than that of HMM and SVM classifiers on every kind of features. A most salience characteristic of Bayesian Networks classifier is the interdependencies of the causal relationships among the component variables, which is also the foundation of naive Bayes' rule. In online sketchy shape recognition, although the features extracted from the strokes can be as interdependent as possible, there are at least time and serial orders among the strokes and people always input certain strokes before others when drawing a specific symbol, which do not perfectly conform to the independence hypothesis and affect the recognition results.

4 Conclusion

In this paper, we develop a method for adaptive online sketchy shape recognition based Bayesian Networks. A sketch is modeled, by means of Bayesian Networks, with the relationships not only between stroke and its neighbouring strokes, but also between a stroke and all of its subsequence. The drawing-style tree is then adopted to capture the users' accustomed drawing styles and simplify the training and recognition of Bayesian Networks classifier. A most salience characteristic of Bayesian Networks classifier for adaptive online sketchy shape recognition is that it can effectively model the intrinsic temporal orders among the strokes. The experiments prove both effectiveness and efficiency of the proposed method.

Acknowledgement

The work described in this paper is supported by the grants from "the National Natural Science Foundation of China" [Grants No. 69903006 and 60373065] and "the Program for New Century Excellent Talents in University of China" [Grant No. NCET-04-04605].

References

1. Zhengxing Sun, Jing Liu, Informal user interface for graphical computing, Lecture Notes in Computer Science, Volume 3784, 2005, Pages 675-682.
2. Landay J. A. and Myers B. A.: Sketching Interfaces: toward more human interface design. IEEE Computer, Vol. 34, No. 3, 2001, page 56-64.

3. Newman M. W., James L., Hong J. I., et al: DENIM: An informal web site design tool inspired by observations of practice. HCI, Vol. 18, 2003, page 259-324.
4. Fonseca M J, Pimentel C.and Jorge J A: CALI - an online scribble recognizer for calligraphic interfaces, AAAI Spring Symposium on Sketch Understanding, AAAI Press (2002), page 51-58.
5. Chris Calhoun, Thomas F S, Tolga Kurtoglu, et al: Recognizing multi-stroke symbols, AAAI Spring Symposium on Sketch Understanding, AAAI Press, 2002, page 15-23.
6. Xiaogang Xu, Zhengxing Sun, et al, An online composite graphics recognition approach based on matching of spatial relation graphs. IIDAR, Vol. 7, No.1, 2004, Pages 44-55.
7. Zhengxing Sun, Wenyin Liu, Binbin Peng, et al, User adaptation for online sketchy shape recognition, Lecture Notes in Computer Science, Volume 3088, 2004, Pages 303-314.
8. Zhengxing Sun, Lisha Zhang and Enyi Tang, An incremental learning algorithm based on SVM for online sketchy shape recognition, Lecture Notes in Computer Science, Volume 3610, 2005, Pages 655-659.
9. Sezgin T. M. and Davis R., HMM-Based Efficient Sketch Recognition, Proceedings of the 10th international conference on IUI, Jan., 2005, San Diego, California, USA.
10. Zhengxing Sun, Wei Jiang and Jianyong Sun , Adaptive Online Multi-Stroke Sketch Recognition based on Hidden Markov Model, Lecture Notes in Artificial Intelligences, Volume 3784, 2005, Pages 948-957..
11. Friedman N, Geiger D and Goldszmidt M, Bayesian network classifiers, Machine learning, Kluwer Academic Publishers, Hingham, USA, Vol. 29, Issue 2-3, 1997, page 131-163.
12. Cho Sung-Jung and Kim Jin H., Bayesian network modeling of Hangul characters for online handwriting recognition, Proceedings of IDAR2003, page 207-211.
13. Alvarado. C and Davis R., Dynamically Constructed Bayesian Networks for Sketch Understanding, Proceedings of IJCAI-05, Edinburgh, Scotland.

Robust Object Tracking Algorithm in Natural Environments

Shi-qiang Hu[1], Guo-zhuang Liang[1], and Zhong-liang Jing[2]

[1] College of Informatics and Electronics, Hebei University of Science and Technology, Shijiazhuang, 050054, China
sqhu@mail.sjtu.edu.cn
[2] Institute of Informatics and Electronics, Shanghai Jiaotong University, Shanghai, 200030, China

Abstract. In order to realize robust visual tracking in natural environments, a novel algorithm based on adaptive appearance model is proposed. The model can adapt to changes in object appearance over time. A mixture of three Gaussian distributions models the value of each pixel. An online Expectation Maximization (EM) algorithm is developed to update the parameters of the Gaussians. The observation model in the particle filter is designed based on the adaptive appearance model. Numerous experimental results demonstrate that our proposed algorithm can track objects well under illumination change, large pose variation, and partial or full occlusion.

Keywords: visual tracking; adaptive appearance model; Expectation Maximization algorithm; particle filter.

1 Introduction

Visual tracking has very important applications in many fields, such as vision-based control, smart surveillance, and visual recognition. One of the key challenges in visual tracking is to handle the appearance variability of an object over time. This variability results from many factors, including variation in object pose or object deformations, partial or full occlusion. Therefore, an important research in visual tracking is to design a robust tracking algorithm and adaptive appearance model in natural environments.

In conventional tracking algorithms, the appearance model is either fixed or rapidly updated. In the first case, the appearance model is extracted from the first frame and kept unchanged during tracking, such as color models in [1, 2], and gray-level template in [3]. In the second case, the appearance model is updated every frame using the tracking results from the previous frames, such as gray-level templates in [4, 5, 6], and color model in [7]. Both kinds of the methods are unavailable. If the appearance model is fixed, the tracking will fail in natural environments such as illumination change, pose variation, or occlusion. However, if the appearance model is updated rapidly, the tracker can easily use wrong appearance information due to partial occlusion or due to the accumulation errors from the previous tracking steps.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 516 – 525, 2006.

In this paper, we propose a robust objects tracking algorithm based on adaptive appearance model. The model can adapt to slowly or rapidly changing appearance. In this model, the value of each pixel over time is considered as a "pixel process" and each "pixel process" is modeled with a mixture of three Gaussian distributions. To adapt to changes in object appearance over time, an online EM algorithm is developed to update the Gaussian parameters. The adaptive appearance model is incorporated in the particle filter. Experimental results in indoor and outdoor video sequences demonstrate the effectiveness and robustness of our proposed algorithm.

2 Adaptive Appearance Model

2.1 Mixture Appearance Model

Due to pose variation or occlusion, a three-Gaussian mixture model is constructed for each pixel to model the distribution of value observed at that pixel over time in this paper. Our mixture model for appearance shows some similarity to the model used in [8] for background. We use A_t to denote the appearance model at time t. Assume the recent history values of a given pixel i, denoted by $Z_{1:t}(i) = \{Z_1(i), \ldots, Z_t(i)\}$, are regarded as a statistical process independent of that for all other pixels, then the probability of the current pixel value $Z_t(i)$ in A_t can be estimated by:

$$p\left(Z_t(i) \mid \Theta_t(i)\right) = \sum_{m=1}^{3} \pi_{m,t}(i)\, p\left(Z_t(i) \mid \theta_{m,t}(i)\right) \text{ and } \sum_{m=1}^{3} \pi_{m,t}(i) = 1 \tag{1}$$

where $\theta_{m,t}(i) = (\mu_{m,t}(i), \sigma_{m,t}(i))$, $\Theta_t(i) = (\pi_{m,t}(i), \theta_{m,t}(i))_{m=1}^{3}$, $\pi_{m,t}(i)$ is the weight of the mth Gaussian at time t, $p\left(Z_t(i) \mid \theta_{m,t}(i)\right)$ is a normal density represented by:

$$p\left(Z_t(i) \mid \theta_{m,t}(i)\right) = N\left(Z_t(i); \mu_{m,t}(i), \sigma^2_{m,t}(i)\right) = \frac{1}{\sqrt{2\pi}\sigma_{m,t}(i)} \exp\left(-\frac{\left(Z_t(i) - \mu_{m,t}(i)\right)^2}{2\sigma^2_{m,t}(i)}\right) \tag{2}$$

2.2 Parameter Online Estimation

We use EM algorithm to estimate $\Theta_t(i)$ parameters. To make the parameters estimation depend more heavily on the most recent observations, define a forgetting factor α. The log-likelihood expression for the observation at given pixel i, $Z_{1:t}(i)$, is given by:

$$L\left(\Theta_t(i) \mid Z_{1:t}(i)\right) = \sum_{k=1}^{t} \alpha^{t-k} \log\left[\sum_{m=1}^{3} \pi_{m,t}(i)\, p\left(Z_k(i) \mid \theta_{m,t}(i)\right)\right] \qquad \alpha \in (0,1] \tag{3}$$

Our aim is to find that $\Theta_t(i)$ which maximizes the log-likelihood function. Similar to standard EM algorithm [9], $\Theta_t(i)$ can be estimated by iterating the following two steps:

Step 1. E-step

$$o_{m,k}(i) = p\left(m \mid Z_k(i), \Theta_{t-1}(i)\right) = \frac{\pi_{m,t-1}(i)\, p\left(Z_k(i) \mid \theta_{m,t-1}(i)\right)}{\sum_{m=1}^{3} \pi_{m,t-1}(i)\, p\left(Z_k(i) \mid \theta_{m,t-1}(i)\right)} \tag{4}$$

Step 2. M-step

$$\pi_{m,t}(i) = \frac{\sum_{k=1}^{t} \alpha^{t-k} o_{m,k}(i)}{\sum_{m=1}^{3} \sum_{k=1}^{t} \alpha^{t-k} o_{m,k}(i)} \tag{5}$$

$$\mu_{m,t}(i) = \frac{\sum_{k=1}^{t} \alpha^{t-k} o_{m,k}(i) Z_k(i)}{\sum_{k=1}^{t} \alpha^{t-k} o_{m,k}(i)} \tag{6}$$

$$\sigma_{m,t}^2(i) = \frac{\sum_{k=1}^{t} \alpha^{t-k} o_{m,k}(i) \left[Z_k(i) - \mu_{m,t}(i)\right]^2}{\sum_{k=1}^{t} \alpha^{t-k} o_{m,k}(i)} \tag{7}$$

In the EM algorithm, computing ownership $o_{m,k}(i)$ would require storing all of the observations from previous times. Considering that the parameters of the Gaussians change slowly, assume $o_{m,t}(i) \approx o_{m,t-1}(i)$, the approximation equations as follows:

$$o_{m,t}(i) = \frac{\pi_{m,t-1}(i)\, p\left(Z_t(i) \mid \theta_{m,t-1}(i)\right)}{p\left(Z_t(i) \mid \Theta_{t-1}(i)\right)} \tag{8}$$

$$\pi_{m,t}(i) = \alpha \pi_{m,t-1}(i) + (1-\alpha) o_{m,t}(i) \tag{9}$$

$$\beta_{m,t}(i) = (1-\alpha) \frac{o_{m,t}(i)}{\pi_{m,t}(i)} \tag{10}$$

$$\mu_{m,t}(i) = \left[1 - \beta_{m,t}(i)\right] \mu_{m,t-1}(i) + \beta_{m,t}(i) Z_t(i) \tag{11}$$

$$\sigma_{m,t}^2(i) = \left[1 - \beta_{m,t}(i)\right] \sigma_{m,t-1}^2(i) + \beta_{m,t}(i) \left[Z_t(i) - \mu_{m,t}(i)\right]^2 \tag{12}$$

2.3 Model Update

Fig. 1 shows that the update procedure. In the initialization step, the distribution of the pixel is represented by the first Gaussian component. If at least one match is found, the parameters will be updated using the current pixel value. If none of the Gaussian distributions match the current pixel value, a reset step will be implemented. It is noted that when the object is occluded, the current update process will stop, and the appearance model is kept unchanged.

Step 1. Initialization

$\pi_{1,0}(i)=1, \pi_{2,0}(i)=\pi_{3,0}(i)=0$; $\mu_{1,0}(i)=T_0(i)$, $\mu_{2,0}(i)=\mu_{3,0}(i)=Inf$;

$\forall m=1,2,3, \sigma_{m,0}^2(i)=\sigma_0^2, C_m(i)=0$

Step 2. For t = 1, …,

(i) For m = 1, 2, 3,

(ii) $$p\left(Z_t(i)\,|\,\theta_{m,t-1}(i)\right)=\begin{cases} N\left(Z_t(i);\mu_{m,t-1}(i),\sigma_{m,t-1}^2(\mathrm{i})\right) & \text{if } \left|Z_t(i)-\mu_{m,t-1}(i)\right|\le T_\sigma\sigma_{m,t-1}(\mathrm{i}) \\ 0 & \text{otherwise} \end{cases}$$

(ii) $$p\left(Z_t(i)\,|\,\Theta_{t-1}(i)\right)=\sum_{m=1}^{3}\pi_{m,t-1}(i)\,p\left(Z_t(i)\,|\,\theta_{m,t-1}(i)\right)$$

(iii) If $p\left(Z_t(i)\,|\,\Theta_{t-1}(i)\right)>0$

For m = 1, 2, 3 // Update step

$C_m(i)$=0 // Once the update condition is met, all the counters are cleared

$$o_{m,t}(i)=\pi_{m,t-1}(i)\,p\left(Z_t(i)\,|\,\theta_{m,t-1}(i)\right)\Big/ p\left(Z_t(i)\,|\,\Theta_{t-1}(i)\right)$$

$$\pi_{m,t}(i)=\alpha\pi_{m,t-1}(i)+(1-\alpha)o_{m,t}(i)$$

If $o_{m,t}(i)>0$

$$\beta_{m,t}(i)=(1-\alpha)o_{m,t}(i)/\pi_{m,t}(i)$$

$$\mu_{m,t}(i)=\left[1-\beta_{m,t}(i)\right]\mu_{m,t-1}(i)+\beta_{m,t}(i)Z_t(i)$$

$$\sigma_{m,t}^2(i)=\left[1-\beta_{m,t}(i)\right]\sigma_{m,t-1}^2(i)+\beta_{m,t}(i)\left[Z_t(i)-\mu_{m,t}(i)\right]^2$$

End If

End For

Else // Reset step

For $m=1,2,3, \pi_{m,t}(i)=\alpha\pi_{m,t-1}(i)$; $k=\arg\min_m\{\pi_{m,t}(i)\}$

If $C_k(i)\le T_c$

$$\pi_{k,t}(i)=\pi_0, \mu_{k,t}(i)=T_0(i), \sigma_{k,t}^2(i)=\sigma_0^2$$

$C_k(i)$ = $C_k(i)$+1 // The counter of the kth Gaussian is incremented by 1

Else

$$\pi_{k,t}(i)=\pi_0, \mu_{k,t}(i)=Z_t(i), \sigma_{k,t}^2(i)=\sigma_0^2$$

$C_k(i)$=0 // The counter of the kth Gaussian is cleared

End If

For m = 1, 2, 3, $\pi_{m,t}(i)=\pi_{m,t}(i)\Big/\sum_{m=1}^{3}\pi_{m,t}(i)$

End If

Fig. 1. The update procedure for the appearance model

3 Visual Tracking with Adaptive Appearance Model

3.1 Dynamic Model

The motion of an object is represented by a parameterized image warp. The warp parameters are defined as $W_t = (u_t, v_t, s_t, r_t)$, where u_t, v_t, s_t, r_t correspond to translation, scale, and rotation respectively.

$$X_t = \begin{pmatrix} 1 & \Delta T & & & 0 & \\ 0 & 1 & & & & \\ & & 1 & \Delta T & & \\ & & 0 & 1 & & \\ 0 & & & & 1 & \\ & & & & & 1 \end{pmatrix} X_{t-1} + \begin{pmatrix} \Delta T^2/2 & & & & 0 & \\ T & & & & & \\ & & \Delta T^2/2 & & & \\ & & T & & & \\ 0 & & & & 1 & \\ & & & & & 1 \end{pmatrix} U_t \tag{13}$$

where $X_t = (u_t, \dot{u}_t, v_t, \dot{v}_t, s_t, r_t)^{\mathrm{T}}$, $\dot{u}_t$ and $\dot{v}_t$ are the velocities of u_t and v_t, Σ is a diagonal matrix whose elements are the corresponding variances of four parameters in W_t, i.e., $\sigma_u^2, \sigma_v^2, \sigma_s^2, \sigma_r^2$.

3.2 Adaptive Observation Model

Given a particle state $X_t^{(j)}$, the candidate object region is determined by the following transformation:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = s_t^{(j)} \begin{pmatrix} \cos(r_t^{(j)}) & -\sin(r_t^{(j)}) \\ \sin(r_t^{(j)}) & \cos(r_t^{(j)}) \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} u_t^{(j)} \\ v_t^{(j)} \end{pmatrix} \tag{14}$$

The observation of the candidate object region can be represented as:

$$\hat{Z}_t^{(j)} = I_t\left(\psi(\mathbf{x}_{t-1}, X_t^{(j)})\right) \tag{15}$$

After obtaining the observation $\hat{Z}_t^{(j)}$ corresponding to the state $X_t^{(j)}$, the observation likelihood $p(Y_t \mid X_t^{(j)})$ is designed as:

$$p\left(Y_t \mid X_t^{(j)}\right) = p\left(\hat{Z}_t^{(j)} \mid X_t^{(j)}\right) = \prod_{i=1}^{d} \left\{ \sum_{m=1}^{3} \pi_{m,t-1}(i) N\left(\hat{Z}_t^{(j)}(i); \mu_{m,t-1}(i), \sigma_{m,t-1}^2(i)\right) \right\} \tag{16}$$

4 Occlusion Handling

In general, outlier pixels cannot be explained by the underlying process and their influences on the estimation process should be decreased. We adopt the Huber's function ρ, which is defined as follows:

$$\rho(\varepsilon) = \begin{cases} c\left(|\varepsilon| + \dfrac{c}{2}\right) & \text{if } |\varepsilon| > c \\ \dfrac{1}{2}\varepsilon^2 & \text{otherwise} \end{cases} \tag{17}$$

At time t = 0.
Step 1. Initialization
For j = 1, …, N, sample $X_0^{(j)}$ from prior distribution $p(X_0)$
Set *Occ_Flag* = 0 to indicate no occlusion
At time t = 1, 2, …
Step 2. Important sampling step
For j = 1, …, N
Propagate $\{X_{t-1}^{(j)}, w_{t-1}^{(j)}\}$ by sampling $\tilde{X}_t^{(j)}$ using the dynamic model
Compute the warped image $\hat{Z}_t^{(j)}$ and $p(\hat{Z}_t^{(j)} \mid \tilde{X}_t^{(j)})$
Update the weight using $w_t^{(j)} \propto p(Y_t \mid \tilde{X}_t^{(j)}) = p(\hat{Z}_t^{(j)} \mid \tilde{X}_t^{(j)})$
End For
For j = 1, …, N, normalize the weight $w_t^{(j)} = w_t^{(j)} / \sum_{j=1}^{N} w_t^{(j)}$
Step 3. Resampling step
Resample $\{\tilde{X}_t^{(j)}, w_t^{(j)}\}_{j=1}^{N}$ to obtain a new particle set $\{X_t^{(j)}, 1/N\}_{j=1}^{N}$
Step 4. State estimation
Estimate the state $\hat{X}_t = \frac{1}{N}\sum_{j=1}^{N} X_t^{(j)}$
Step 5. Appearance model update
Compute the warped image Z_t using the state $\hat{X}_t$
Set *Occ_Flag* according to the number of the outlier pixels in Z_t
If *Occ_Flag* = 0
Update the appearance model A_{t-1} using Z_t
Else
A_{t-1}=A_t
End If

Fig. 2. The tracking algorithm with occlusion handling

Where the constant c is a scale parameter, the function ρ has the following expression:

$$\rho\left(\frac{Z_t(i)-\mu_{m,t-1}(i)}{\sigma_{m,t-1}(i)}\right)=\begin{cases} T_\sigma\left(\left|\frac{Z_t(i)-\mu_{m,t-1}(i)}{\sigma_{m,t-1}(i)}\right|-\frac{T_\sigma}{2}\right) & \text{if } \forall m=1,2,3, \left|\frac{Z_t(i)-\mu_{m,t-1}(i)}{\sigma_{m,t-1}(i)}\right|>T_\sigma \\ \frac{1}{2}\left(\frac{Z_t(i)-\mu_{m,t-1}(i)}{\sigma_{m,t-1}(i)}\right)^2 & \text{otherwise} \end{cases} \tag{18}$$

Where $Z_t(i)$, T_σ, $\mu_{m,t-1}(i)$, and $\sigma_{m,t-1}(i)$ are the same as those in the previous sections.

The proposed objects tracking algorithm with occlusion handling is shown in Fig. 2.

5 Experiments and Results Analysis

In this section, we test our algorithm on many real-word video sequences [11]. These sequences contain many difficult tracking conditions in natural environments, including illumination change, large pose variation and partial or full occlusion. Some parameters are set as: $\alpha = 0.85$, $\sigma_0 = 20$, $\pi_0 = 0.1$, $T_\sigma = 3$, $T_c = 5$. The tracking result is shown with a white bounding box. Table 1 summarizes some statistics about these video sequences and the appearance model size used.

Table 1. Comparison of tracking results with different appearance models

Video		Video1	Video2	Video3	Video4
# of frames		500	1540	1144	550
Frame size		96x128	240x320	240x320	240x320
A_t size		43x37	40x32	45x32	36x24
Lost tracks	Non-adaptive	136	254	318	240
	Adaptive	0	0	0	0

5.1 Tracking a Head

The video sequence shown in Fig. 3 is a woman head moving in an office environment. The tracking results show that the adaptive tracking algorithm can track the object successfully throughout the sequence.

(a) Tracking using the fixed appearance model

(b) Tracking using the adaptive appearance model

Fig. 3. Tracking results with head rotation and cluttered backgrounds

The RMSE for the non-adaptive tracking algorithm is computed as:

$$RMSE_{Fixed}(t) = \sqrt{\frac{1}{d}\sum_{i=1}^{d}\left[Z_t(i) - T_0(i)\right]^2} \tag{19}$$

The RMSE for the adaptive tracking algorithm is computed as:

$$RMSE_{Adaptive}(t) = \sqrt{\frac{1}{d}\sum_{i=1}^{d}\left[Z_t(i) - \sum_{m=1}^{3}\pi_{m,t}(i)\mu_{m,t}(i) \right]^2} \tag{20}$$

The RMSE curves indicate that the adaptive appearance model invokes smaller.

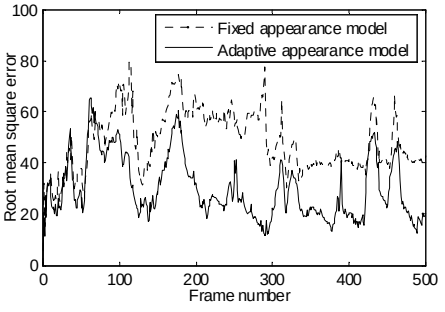


Fig. 4. The RMSE for two kinds of the trackers

5.2 Tracking Face Under Illumination Change

Fig. 5 shows the results of tracking a man face under illumination change. The results show that the adaptive tracker can track the face robustly, while the non-adaptive tracker fails when there is a large illumination change.

(a) Tracking using the fixed appearance model

(b) Tracking using the adaptive appearance model

Fig. 5. Tracking results under illumination change

5.3 Tracking Head Under Full Occlusion

The sequence shown in Fig. 6 is a man head fully occluded twice by a tree. The results are obtained by incorporating the occlusion analysis in the adaptive tracker. It can be seen from the results that when the object become visible again after a period of occlusion, our tracker can succeed to track it. Fig. 7 shows that the occlusion handling can succeed to judge the time when occlusion happens.

Fig. 7. Tracking results under full occlusion

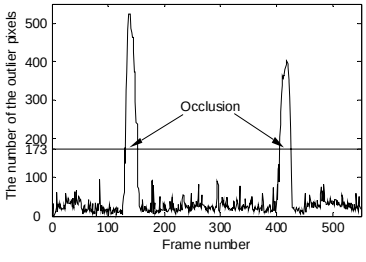


Fig. 8. The number of the outlier pixels vs. frame number

6 Conclusions

We have presented an algorithm for tracking objects in natural environments using adaptive appearance model. An online EM algorithm was developed to handle appearance changes between frames. The adaptive appearance model was incorporated in the particle filter. Occlusion analysis was implemented using robust statistics. Demonstrated by five challenge video sequences, the algorithm can robustly track an object in many complex situations.

Future work will need to address the better model characterizing object maneuvers. In addition, more optimal algorithm suitable for practical application will be developed.

Acknowledgements

This work was jointly supported by Hebei Natural Science Foundation (F2006000343) and Hebei PH.D Discipline Special Foundation (B2004510).

References

1. D. Comaniciu, V. Ramesh, and P. Meer, Real-time tracking of non-rigid objects using mean shift. Proc. IEEE Conf. Computer Vision and Pattern Recognition, (2000)142-149.
2. T.L. Liu and H.T. Chen, Real-time tracking using trust-region methods. IEEE Trans. Pattern Analysis and Machine Intelligence, (2004), 397-402.

3. G.D. Hager and P.N. Belhum, Efficient region tracking with parametric models of geometry and illumination. IEEE Trans. Pattern Analysis and Machine Intelligence, (1998), 1025-1039.
4. H.T. Nguyen, M. Worring, and R. Boomgaard, Occlusion robust adaptive template tracking. Proc. Int. Conf. Computer Vision, (2001) , 678-683.
5. N. Papaniko, P. Khosla, and T. Kanade, Visual tracking of a moving target by a camera mounted on a robot: a combination of control and vision. IEEE Trans. Robotics and Automation, (1993), 14-35.
6. H. Sidenblad, M.Black, and D Fleet, Stochastic tracking of 3D human figures using 2D image motion. Proc. European Conf. Computer Vision, (2000), 702-718.
7. Y. Wu and T. S. Huang, Color tracking by transductive learning. Proc. IEEE Conf. Computer Vision and Pattern Recognition, (2000),133-138.
8. D. Lee, Effective Gaussian mixture learning for video background subtraction. IEEE Trans. Pattern Analysis and Machine Intelligence, (2005), 827-832.
9. A.P. Dempster, N.M. Laird, and D.B. Rubin, Maximum likelihood from incomplete data via the EM Algorithm. J. Royal. Statistical. Soc. (1977), 1-38.
10. M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian bayesian tracking. IEEE Trans. Signal Processing, (2002), 174-188.
11. http://vision.stanford.edu/~birch, http://www.cs.toronto.edu/~dross/ivt,
12. http://www.cs.toronto.edu/vis/projects/dudekfaceSequence.html.

An Image Retrieval Method on Color Primitive Co-occurrence Matrix

HengBo Zhang[1,2], ZongYing Ou[1], and Guanhua Li[1]

[1] Key Laboratory for Precision and Non-traditional Machining Technology of Ministry of Education, Dalian University of Technology, P.R. China, Dalian 116024
[2] Res. Inst. of CG & Image Processing,Dalian Nationalities University, P.R. China, Dalian 11660
zhanghengbo@dlnu.edu.cn

Abstract. The paper presents and realizes a new image retrieval method based on the combination of the color connected area information with the texture features. The image is firstly divided into several parts and the color connected areas in the image is computed, then, the primitive co-occurrence matrix of the four color components corresponded with the connected area of each color is extracted. Lastly the image retrieval on the basis of the content is realized by using of the feature similar function which is designed according to these features.

1 Introduction

This paper present and realize the new image retrieval method that combining the distribution of color and texture of image. Based on color connected region, the method combines the information of the color constitution and its distribution in the image and gets each of features of the co-occurrence matrix relevant to the each component of color. Then, it realizes the Content Based Image Retrieval on the basis of the image similitude measure function of this feature. The results of the experiment in this paper demonstrate that the effect of the image retrieve method is well.

2 Grey-Primitive Co-occurrence Matrixes

2.1 Primitive Arrays

If there exists a image $f(i,j)$, i=0,1,2,… ,Lx-1; j =1,2,… ,Ly-1, we will discuss the constitution of the primitive array in the forms of the primitive like Fig.1 (a) and (b). The value of primitive is measured by moment of neighborhood pixels coping with the pixel. Except for the boundary pixels, calculating the moment of each pixel that 4 neighborhood pixels grey level cope with the pixel:

$$m(i,j) = f(i-1,j) + f(i,j+1) + f(i+1,j) + (i,j-1) \quad (i = 1,2,...,Lx-2;\ j = 1,2,...,Ly-2) \tag{1}$$

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 526 – 529, 2006.

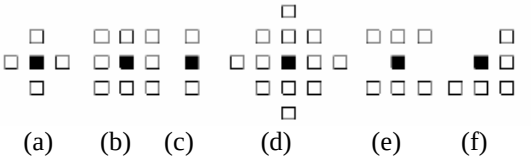


Fig. 1. The structure of primitive

2.2 Grey Primitive Co-occurrence Matrix

Let $Lgpx=f_{normalize}$, and let $Lgpy=m_{normalize}$. The following *GP(i,j)* is the define grey primitive co-occurrence matrix.

The matrix of *GP(i, j)*, i= 0, 1, 2, … , *Lgpx*; j = 0, 1, 2,… , *Lgpy*, whose dimension is (*Lgpx*+1)×(*Lgpy*+1) consists of grey matrix and primitive matrix. The value of element of marix *GP(i, j)* is the number of points which the grey equal to i and the primitive equal to j.

Take *GP*(5,8)=16 as an example, in the grey-primitive co-occurrence matrix of a image, the number of points is 16. the grey of this kind point equals to 5, and the primitive equals to 8.

3 Set of Color-Connected Regions[4]

3.1 Color Connected

A color image *T* is divided into *N*×*N* sub images. And, calc is calculated according to ulate the dominant color *C(i,j)* of every sub image *T(i,j)* ($1\le i\le N$, $1\le j\le N$) according to the mosaic algorithmic of extracting dominant colors. If two arbitrary sub images that are called *T(i,j)* and *T(k,l)*(i-k=1 and j-l=1) conform to 8-connected regulation and the corresponding dominant colors are interrelated, the sub images *T(i,j)* and *T(k,l)* are color-connected.

3.2 Set of Color-Connected Regions

According to 8-connected regulation, compose all the sub images as a set of color-connecting regions $S=\{R_i\}$ ($1\le i\le M$), which conform to the conditions as follows (*M* is the number of color-connecting regions):

(a) $\bigcup_{i=1}^{M} R_i = T$

(b) $R_i \cap R_j = \Phi(i \ne j, 1 \le i, j \le M)$

(c) R_i is a set of color-connected regions

(d) R_i and R_j are color-unconnected

4 Algorithm of Retrieving Image

4.1 Distill Texture Feature of Color Primitive Co-occurrence Matrix

Distill each feather of grey primitive co-occurrence matrix from each region R_i ($1 \le i \le M$) of the set S of color-connecting regions, which is in corresponding to the range of R_i, then, compose them together to form the texture signature of color primitive co-occurrence matrix of this image.

Step1, Divide color image of T into $N \times N$ sub images and form the set $S=\{R_i\}(1 \le i \le M)$ of color-connected regions according to the color-connected regulation which is mentioned above.

Step2, Extract matrix information of four color components of R, G, H and I which belong among image of T. And the maximal gradation of each component equals to D=8.

Step3, As is mentioned in 2.2, we ought to calculate primitive co-occurrence matrix which has been normalized of four color components of R, G, H and I. for each color connecting region R_i ($1 \le i \le M$). Then, for each co-occurrence matrix, we should calculate statistical characteristic quantities such as E(Energy), I(Inertia), S(Entropy), H(Evenness):[5]

The texture feature vector $F_i=[F_{Ri},F_{Gi},F_{Hi},F_{Ii}] = \{f_{i1}, \ldots, f_{i16}\}$, which is 16 dimensions consists of characteristic quantities which are mentioned above .

Step4, There is a hypothesis that the number of sub images of each color-connecting region equals to B. And, calculate the texture feature vector $F= [F_R, F_G, F_H, F_I]= \{f_1,\ldots,f_{16}\}$.:

Step5, over.

4.2 Similarity Measurement Function

Assuming image A and B, the color primitive matrix characteristic respectively is:

$$F_A=[F_{RA},F_{GA},F_{HA},F_{IA}]=\{f_{A1},\ldots,f_{A16}\}, \quad F_B=[F_{RB},F_{GB},F_{HB},F_{IB}]=\{f_{B1},\ldots,f_{B16}\}$$

Four parts in the above-mentioned characteristic vector represent the different aspect of the characteristic respectively. The similarity measurement function is constituted by four items, showing such as the following formula:

$$\begin{aligned} D(A,B) = {} & w_1 \cdot D_E(F_{RA},F_{RB}) + w_2 \cdot D_E(F_{GA},F_{GB}) \\ & + w_3 \cdot D_E(F_{HA},F_{HB}) + w_4 \cdot |F_{GA} - F_{GB}| \end{aligned} \tag{2}$$

5 Analysis of the Experiment Results

This paper has set up a database of 500 images which are collected from web. It includes nature scenes, animals, flowers, food, cars and so on, which are all 24-bit true-color image. We can find this algorithm is effect when rotated.

Table 1 listed the comparison of retrieval with grey co-occurrence matrix, grey primitive co-occurrence matrix and color primitive co-occurrence matrix.

Table 1. Retrieval result comparison

Co-occurrence matrix	Template content	Correctly retrieved sample number	Erringly retrieved sample number	Accurate rate(%)
Grey	Food	29	21	58
	Snow	31	19	62
Color primitive	food	44	6	88
	snow	46	4	92

6 Conclusion

As referring to the color structure and distributing information, the features which are extracted in this way can reflect both texture relation and color structure. It has also built the corresponding relationship with people's perception. The experiment has turned out that this method is better than the method of block grey co-occurrence matrix and the method of color co-occurrence matrix on whole image. This method has better retrieval effect and is applicable to contend-based image retrieval.

References

[1] Koskela M, Laaksonen J, Oja E. Comparison of techniques for content-based image retrieval[A]. Proc of 12th Scandinavian Conference on Image Analysis[C]. Bergen, Norway: Norwegian Society for Image Processing and Pattern Recognition, 2001.579-586.

[2] Manjunath B S, Ma W Y. Texture features for browsing and retrieval of image data[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(8): 837-842.

[3] Wang Liangshen, OU Zongying. Image Texture Analysis by Grey-Primitive Co-occurrence Matrix[J]. Computer Engineering, 2004, 30(23):19-21

[4] Conners R W, Hariow C A. A theoretical comparison of texture algorithms[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1980, 2(3):204-222.

[5] YangYubin, ChenShi-fu, LinHui. A Novel Image Retrieval Method Using Texture Features Based on Color Connected Regions [J]. Acta Electronica Sinica, 2005, 33(1): 58-62

A Modified Adaptive Chaotic Binary Ant System and Its Application in Chemical Process Fault Diagnosis

Ling Wang and Jinshou Yu

Research Institution of Automation, East China University of Science & Technology
200237, Shanghai, China
shwl_1212@163.com

Abstract. Fault diagnosis is a small sample problem as fault data are absent in the real production process. To tackle it, Support Vector Machines (SVM) is adopted to diagnose the chemical process steady faults in this paper. Considering the high data dimensionality in the large-scaled chemical industry seriously spoil classification capability of SVM, a modified adaptive chaotic binary ant system (ACBAS) is proposed and combined with SVM for fault feature selection to remove the irrelevant variables and ensure SVM classifying correctly. Simulation results and comparisons of Tennessee Eastman Process show the developed ACBAS can find the essential fault feature variables effectively and exactly, and the SVM fault diagnosis method combined with ACBAS-based feature selection greatly improve the diagnosing performance as unnecessary variables are eliminated properly.

1 Introduction

Fault diagnosis, which can provide early warning for process upset and reduce loss, plays an important role in industry process. But it is a challenge to apply fault diagnosis to modern complex and large-scaled chemical process because there are large amounts of variables with noise monitored. When all collected variables are taken as the inputs to discriminate faults, high dimension of data greatly reduces the performance of diagnosing because of the disturbances of too much irrelevant variables and noises, and badly spoil the real-time capability due to the increased complexity of computation. Especially, fault diagnosis is a small samples problem as the fault data are deficient in the really production process, which make some classification algorithms not suitable for fault diagnosis.

In this paper, we use the SVM as classifier to diagnose the chemical process steady faults as it can tackle the small sample problem. As mentioned above, high dimension of data will impair the classification capability of SVM. So it is essential to preprocess the sampled data to reduce the data dimension.

There have been several approaches to preprocess data developed, applied and widely researched in fault diagnosis applications, such as Principal Component Analysis (PCA) [1] and Kernel PCA [2], which are well-known methods for feature

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 530 – 539, 2006.

extraction. The large sets of process monitored variables data in the chemical industry are often highly correlated and as a result are generally good candidates for PCA and KPCA. As feature extraction methods can produce lower-dimensional representations of the data which better generalize to data independent of the training set than using the entire dimensionality of the observation space, and therefore, improve the proficiency of detecting and diagnosing fault.

But the extracted information by feature extraction methods is not directly related to the objective of fault diagnosis. So the number of retained components maybe is still large in order to contain enough information for diagnosing. Sometimes, even worse, the extracted data is not exactly acceptable for fault diagnosis because the resulting lower dimensional space may contain little of the required faults information, which makes the feature extraction invalid for fault diagnosis in some conditions.

To make up for this shortage, feature selection method was proposed as an alternative to preprocess the collected data [3]. In the process of fault diagnosis, feature selection is operated to directly search and find the essential fault variables. Only the selected variables will be retained and used as inputs for fault diagnosis. As the irrelative variables are all removed, the real-time capability and correct rates of fault diagnosis will be greatly improved. In order to find the fault feature variables effectively and properly, a modified adaptive chaotic binary ant system is proposed and combined with SVM to realize fault feature selection in the paper.

The reminder of the paper is organized as follows. Section 2 presents the ACBAS algorithm and the feature selection method based on ACBAS combined with SVM in detail. The setup of simulations is introduced in Section 3. Section 4 describes the performances and comparisons of developed ACBAS algorithm and fault diagnosis method based on feature selection. Section 5 concludes the results of simulations.

2 Theory

2.1 Ant Colony Optimization (ACO)

ACO algorithm inspired by colonies of real ants has been successful employed to solve various optimization problems. It is an evolutionary approach where several generations of artificial agents in a cooperative way search for good solutions. Agents are initially randomly generated on nodes, and then stochastically move from a start node to feasible neighbor nodes. Agents collect and store information in pheromone trails during the process of finding feasible solutions. Agents can online release pheromones while building solutions. In addition, the pheromones will be evaporated in the search process to avoid local convergence and to explore more search spaces. Thereafter, additional pheromone is deposited to update pheromone trail offline so as to bias the search process in favor of the currently path.

The pseudo code of the classical ACO algorithm [4] can be described as:

```
Procedure: Ant colony optimization
Begin
   While (ACO has not been stopped) do
      Agents_generation_and_activity();
```

```
        Pheromone_evaporation();
        Daemon_actions();
      End;
  End;
```

More details on ACO algorithm can be found in [5].

2.2 Chaotic Sequences

Recently chaotic sequences have been adopted instead of random ones and the improvements of algorithm performance have been shown in many applications such as neural networks [6], [7] and optimization algorithms. The choice of chaotic sequences is justified theoretically by their unpredictability.

With regard to optimization algorithms, chaotic sequence was introduced and used in the genetic algorithm (GA) [8], [9], evolutionary algorithm (EA) [10] and ACO [11], and the results showed the enhancing of exploitation capability.

In this paper, chaotic sequence has been used in the all the phase of ACO algorithm, i.e., in the generation of the initial population and during the execution of the exploiting and updating pheromone trails.

Here we use the logistic map to produce the chaotic sequences. The logistic equation is defined as follows.

$$x_{n+1} = \mu \cdot x_n \cdot (1 - x_n) \quad 0 < x_0 < 1 . \tag{1}$$

where μ is the control parameter, x is a variable and $n=0,1,2,\ldots$ The logistic equation is chaotic when μ =4 and $x_0 \notin \{0.25, 0.5, 0.75\}$.

2.3 Support Vector Machines

SVM is a relatively new class of machine learning techniques introduced by Vapnik [12]. Unlike most of traditional methods that implement Empirical Risk Minimization Principal, it implements the Structural Risk Minimization Principal by seeking to minimize an upper bound of the generalization error instead of minimizing the training error. This eventually results in better generalization performance than other traditional methods. As the solution is determined only by support vectors that are a subset of all training data points, it has a sparse representation and entails a fast test speed. Because of the remarkable characteristics of SVM, such as good generalization performance, the absence of local minima and the sparse representation of solution, it has been popularly used in the fault diagnosis applications [13-15], but only a few on chemical processes [16].

In this paper, SVM is adopted as fault classifier and combined with the proposed ACBAS for feature selection.

2.4 Fitness Function

To estimate and guide the feature searching of ACBAS, it is essential to measure the performances of fault features selected by each ant. So a pre-defined fitness function is applied to evaluate the fitness. Usually, the fitness function can be simply defined

as the correct classification rate. As irrelative variables are useless and cost much computing time, the fitness function is modified as Eq. (2) to remove the unnecessary variables in this paper.

$$f(x_{id}) = f(id) - p \times \frac{m_c}{m_{all}}. \tag{2}$$

where $f(x_{id})$ means the modified fitness function, $f(id)$ represents the correct fault classification rate, m_c is the number of variables chosen by the ant while m_{all} is the dimension of data samples, p is an important parameter which balances the maximum correct classification rate and the number of retained variables.

2.5 Feature Selection Based on ACBAS

In the proposed ACBAS-based feature selection algorithm, features are represented as graph-ere nodes, with the edges between them denoting the choice of the next feature. Then the fault feature selection is transformed into an ant traversal through the graph. The built solutions of ants are represented as the binary sequences where a bit "1" denotes the corresponding feature is selected and a bit "0" means its corresponding feature is eliminated [17].

The main steps of feature selection based on ACBAS are as follows:

1. Represent all features as graph-ere nodes and initialize the parameters of ACBAS algorithm.
2. If termination criteria is reached then terminate, else go to the next step.
3. Generate ants and visit all features from a chaotic beginning node to build solutions completely according the following rules:

   First of all, produce a chaotic number p_{chaso} by logistic map and compare with the exploiting probability *pe*.

   If p_{chaos} is greater than or equal to the parameter *pe*, whether the feature *j* is selected or not by ant *i* is decided by the two probability parameters $P_{j,0}$, $P_{j,1}$..

$$Solution_{i,j} = \begin{cases} 0 & if \ P_{j,1} \le P_{j,0} \\ 1 & if \ P_{j,1} > P_{j,0} \end{cases} \tag{3}$$

$$\begin{aligned} P_{j,1} &= [\tau_{j,1}]^{\alpha} \cdot [d_i + 1]^{\beta} \\ P_{j,0} &= [\tau_{j,0}]^{\alpha} \cdot [d_i]^{\beta} . \end{aligned} \tag{4}$$

$$d_i = \sum_{k=1}^{j} Solution_{i,k,1} \tag{5}$$

   where d_i is the sum of selected features by ant *i* now. α and β are two parameters that determine the relative influence of the pheromone trail and the heuristic information, i.e., the number of selected features.

   If p_{chaos} is less than *pe*, whether the feature j is chosen is determined by the threshold function:

$$\eta_c(x) = \begin{cases} 0 & c < \eta_0 \\ 1 & c \ge \eta_0 \end{cases} \tag{6}$$

where c is another chaotic number and η_0 is a constant, for instance, 0.5. In our work, we increase η_0 as the heuristic information to speed searching.

4. Calculate the fitness of each ant after all ant agents completing building solutions.
5. The best found solution is reserved and updated. The iteration-best solution (or best found solution sometimes) is chosen to update the pheromones intensities as:

$$\begin{aligned} \tau_{i,0} &= \rho \cdot \tau_{i,0} + (1-\rho) \cdot F / Q \quad \text{if feature } i \text{ not selected} \\ \tau_{i,1} &= \rho \cdot \tau_{i,1} + (1-\rho) \cdot F / Q \quad \text{if feature } i \text{ selected} \end{aligned} \tag{7}$$

where ρ is the evaporation rate to avoid unlimited accumulation of pheromone, F is the fitness of the corresponding ant and Q is an constant.

In the process of ant searching, the exploiting probability *pe* is a very important parameter. A small *pe* can speed convergence of ACO, but it is also likely to stagnate in the local optimal. So a self-adaptive *pe* as Eq. (8) is adopted in the proposed ACBAS algorithm to improve its global searching capability,

$$pe = 0.2 + 0.2 \cdot Ite_c / G_{max} \tag{8}$$

In Eq. (8), Ite_c represents the current generation number and G_{max} is the number of the maximum iterative generations.

3 Experimental Simulation Setup

3.1 Tennessee Eastman Process

The Tennessee Eastman is a well-known benchmark chemical process, which was firstly introduced by Downs and Vogel [18]. The TEP provides a realistic industrial process for evaluating process control and monitoring methods. Now, the TEP has been widely used for the process monitoring community as a source of data for comparing various approaches [19].

The TEP simulator, coded in Matlab, was used to generate normal data and fault data. The pre-defined fault 4, fault 6 and fault 7 with stable operating conditions before and after the faults occur, are researched in this paper.

3.2 Data Sampling

The first simulation ran 25 hours and sampled per 3 minutes to generate500 observations under the normal operating conditions. The next three simulations also ran 25 hours and each of them corresponded to three different faults mentioned above. These three simulations started without faults, and the faults were introduced 1 simulation hour later into the run. So the total number of observations generated for each run was n=500, but only 480 observations were collected after the introduction of the fault. Each observation contains 52 observation variables. As SVM is fit to limited samples, only 30 training data samples and 60 data validation samples data were chosen

randomly from each simulation run and used for fault feature variables selection. Another 4 simulations ran to generate test data to evaluate the performances of fault diagnosis. Like the former 4 simulations, one simulation ran in the normal condition, and the other 3 simulations ran corresponding to fault 4, 6 and 7. But the simulation time for each run was 48 hours. Each simulation started with no faults, and the faults were introduced 8 simulation hours later into the run. The total number of observations generated for each run was n=960.

To make the simulation results comparable, the data used for experiment come from http://brahms.scs.uiuc.edu.

4 Simulation Results

4.1 Fault Feature Selection

In the fault 4 case, all variables remain steady except the 51-th variable induced a change when the fault occurred. So the fault feature variable of fault 4 is variable 51. Fault 6 involves a step changes in variable 1and variable 44, and the other variables are all bothered. The change of variable 1 and variable 44 are so remarkable that any one of them can be taken as fault feature. Variable 45 has a noticeable step change when the fault 7 is introduced into the process. And affected by it, other 34 variables deviate significantly from their normal operation behavior and go aback to normal values later by the control of closed loop.

The experiments were carried out within a Matlab 6.5 enviroment, running on a PC powered by a Pentium IV 2.4GHz and 256MB RAM. To evaluate and compare the developed algorithm, fault feature selection of three faults all ran 20 times using ACBAS algorithm, chaotic binary ant system (CBAS) with *pe=0.2, 0.4*and ACO with the best parameters reported in [18]. All algorithms ran with the same value of parameters except *pe*, set as $N=52, \alpha=1, \beta=0$, $\rho=0.1$, $Q=10$, $G_m=50$, and the ant trails were initialized by logistic map chaotic sequences.

The features selected of 3 faults by all algorithms are shown in Fig. 1–3.

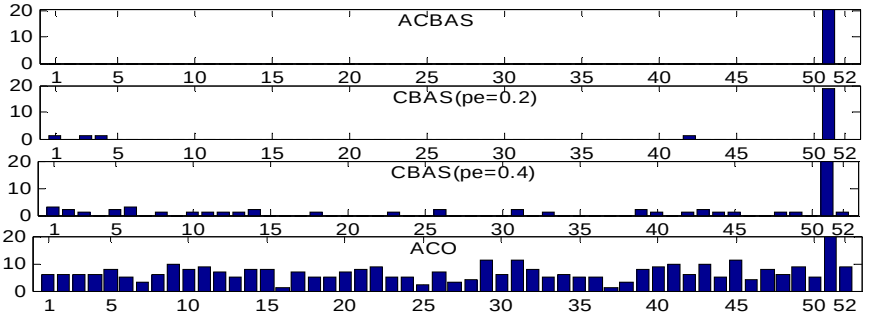


Fig. 1. Fault features of fault 4 selected by all algorithms

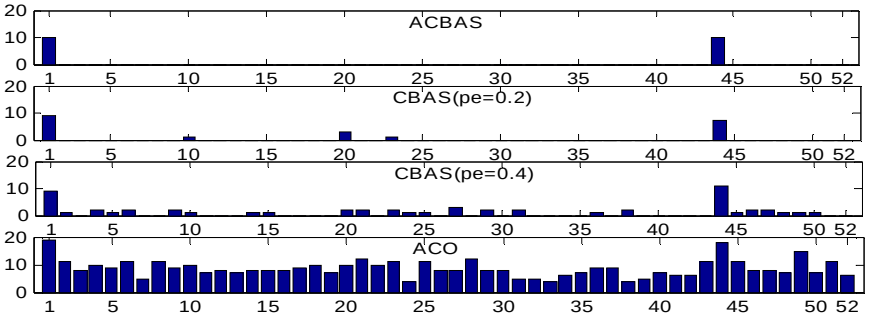


Fig. 2. Fault features of fault 6 selected by all algorithms

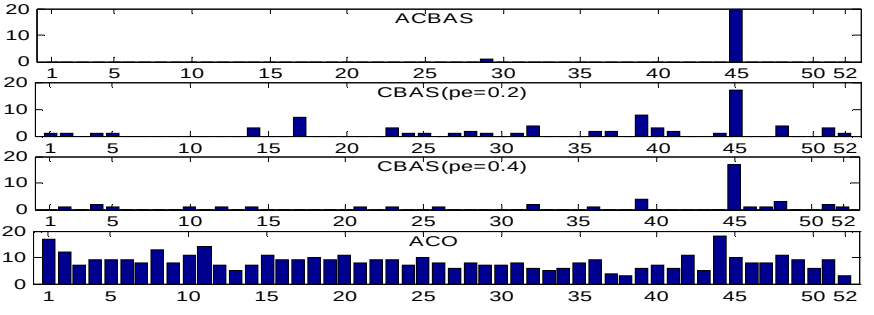


Fig. 3. Fault features of fault 7 selected by all algorithms

And the average fitness changes of all algorithms are figured in Fig. 4–6.

As shown in Fig. 1–3, the introductions of modified heuristic strategy and chaotic sequence greatly improve the initial solutions and the searching ability of ACO algorithm. Ant system with a small *pe* can speed convergence and achieve a better solution while a high value *pe* may make ant system escape from local optimal more effectively as the fitness changes of ant system with different *pe* displayed in Fig. 4–6. According to the results shown in the figures, it is obvious that the proposed ACBAS algorithm can effectively find the fault features prior to classical ACO or the chaotic ant system.

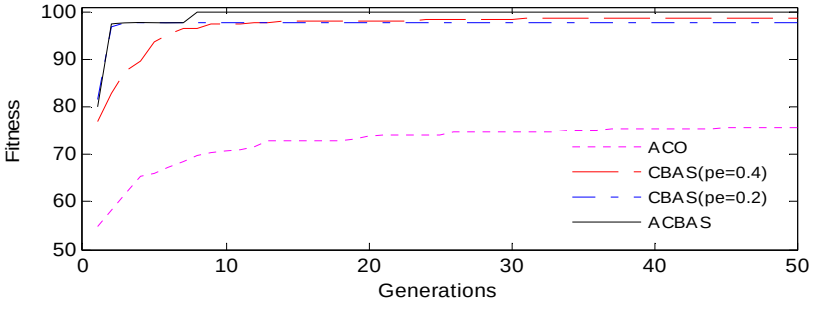


Fig. 4. Fault 4 average fitness changes of all algorithms

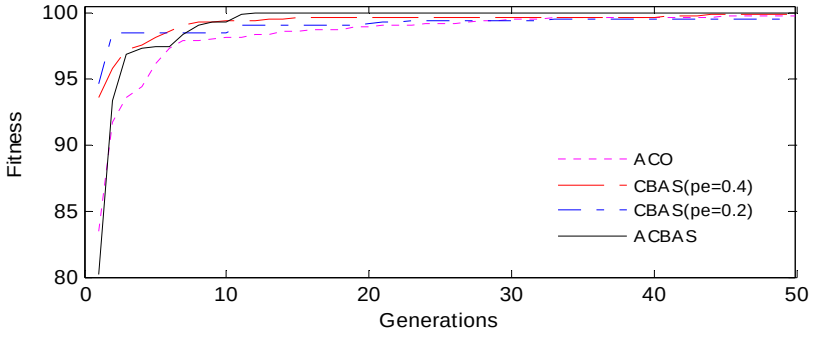


Fig. 5. Fault 6 average fitness changes of all algorithms

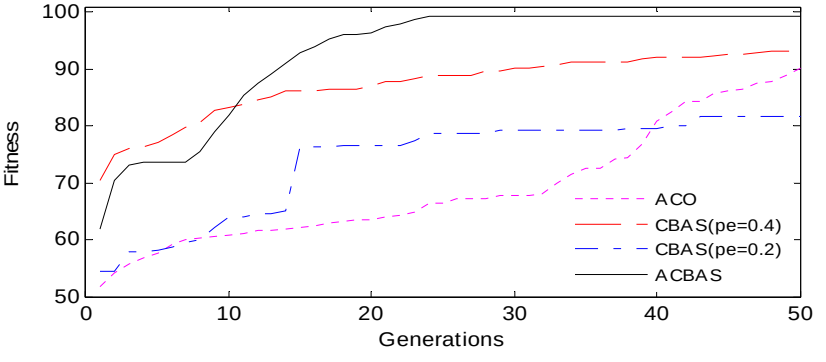


Fig. 6. Fault 7 average fitness changes of all algorithms

4.2 Fault Diagnosis Base on SVM with Fault Feature Selection

Table 1 presents the results of fault diagnosis based on SVM with fault feature selection (FFS). To give a comparison, fault diagnosis methods based on SVM with the all collected data, data extracted by PCA and KPCA were tested with the same data set. The results of these three fault diagnosis methods are given in the Table 1 as well. As the focus of this paper is to develop an effective algorithm to select fault feature and evaluate the fault diagnosis method based on SVM with feature selection in chemical process industrial, the details of whole fault diagnosis strategy are not described here.

Table 1. The correct diagnosing rates of SVM with all variables, PCA, KPCA and feature selection based on ACBAS

	All variables	PCA	KPCA	FFS
Fault 1	61.3%	68.2%	74.5%	100%
Fault 2	99.4%	99.5%	99.9%	100%
Fault 3	51.8%	59.7%	71.2%	100%

5 Results and Discussion

In this paper, we adopted SVM to diagnose the steady faults in chemical process and proposed a modified adaptive chaotic binary ant system algorithm for fault feature selection. The simulation results show the developed ACBAS algorithm has better searching capability and converge more quickly than classical ACO and the chaotic ant system as the adaptive exploiting probability and chaotic sequence are introduced. SVM based on feature selection works perfectly in its applications to diagnosing the chemical process steady faults, as it is suitable for the small sample problem and with good generalization capability. These excellent characteristics make the proposed SVM with feature selection method based on ACBAS algorithm noticeable and attractive in the chemical process fault diagnosis applications.

References

1. Wang S.W., Xiao F.: Detection and diagnosis of AHU sensor faults using principal component analysis method. Energy Conversion and Management. 45 (2004) 2667-2686
2. Lee, J.M., Yoo, C.K., Lee, I.B.: Fault detection of batch processes using multiway kernel principal component analysis Computers and Chemical Engineering. 28 (2004) 1837-1847
3. Chiang, L.H., Pell, R.J.: Genetic algorithms combined with discriminant analysis for key variable identification. Journal of Process Control. 14 (2004) 143-155
4. Bonabeau, E., Dorigo, M., Theraulaz, G.: Swarm Intelligence from Natural to Artificial System. Oxford University Press, Oxford. (1999)
5. Marco D., Christian, B.: Ant colony optimization theory: A survey. Theoretical Computer Science. 344 (2005) 243-278
6. Wang, L, Smit, K.: On chaotic simulated annealing. IEEE Trans. Neural Networks. 9 (1998) 716-718
7. Chen, L., Aihara, K.: Global searching ability of chaotic neural networks. IEEE Trans. Circuit Syst. I. 46 (1999) 974-993
8. Determan, J., Foster, J.A.: Using chaos in genetic algorithm. in: Proceedings of the 1999 Congress on Evolutionary Computation, Vol. 3, Piscataway, NJ:IEEE Press, (1999) 2094-2101
9. Liu, J., Cai, Z.X., Liu, J.Q.: Premature convergence in genetic algorithm: Analysis and prevention based on chaos operator. In: Proc. 3rd World Congress Intelligence Control Automation, Hefei, China. (2000) 495-499
10. Riccardo C., Fortuna, L., Stefano, F., Maria G.X.: Chaotic sequences to improve the performance of evolutionary algorithms. IEEE Transactions on evolutionary computation. 7 (2003) 289-304
11. Flavio, C., Fortuna, L., Mattia, F., Luca, P.: Chaotic sequences in ACO algorithms. In: Proceedings of 2004 IEEE International Symposium on Circuits and Systems. Vol. 4, Piscataway: Institute of Electrical and Electronics Engineers Inc., (2004) 513-516.
12. Vladimir, N.V.: The Nature of Statistical Learning Theory. Springer, New York. (1995)
13. Ge, M, Du, R.: Fault diagnosis using support vector machine with an application in sheet metal stamping operations, Mechanical Systems and Signal Processing. 18 (2004) 143-159
14. Samanta, B., Al-Balushi, K.R.: Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. Engineering Applications of Artificial Intelligence. 16 (2003) 657-665

15. Samanta, B.: Gear fault detection using artificial neural networks and support vector machines with genetic algorithms, Mechanical Systems and Signal Processing. 18 (2004) 625-644
16. Chiang, L.H., Kotanchek, M.E., Kordon, A.K.: Fault diagnosis based on Fisher discriminant analysis and support vector machines Computers and Chemical Engineering. 28 (2004) 1389-1401
17. Chun, K. Z., Hong, H.: Feature selection using the hybrid of ant colony optimization and mutual information for the forecaster. in: Proceedings of Fourth International Conference on Machine Learning and Cybernetics. (2005) 1728-1732
18. Downs J. H., Vogel E.F.: A plant-wide industrial process control problem, Comput. Chem. Eng. 17 (1993) 245-255
19. Chiang, L.H., Russell, E.L., Braatz, R.D..: Fault Detection and Diagnosis in Industrial Systems. Springer-Verlag Berlin Heidelberg London (2001)

Image Context-Driven Eye Location Using the Hybrid Network of k-Means and RBF

Eun Jin Koh and Phill Kyu Rhee

Department of computer science & Engineering Inha University
Biometric Engineering Research Center Young-Hyun Dong, Incheon, Korea
supaguri@im.inha.ac.kr, phrhee@inha.ac.kr

Abstract. In this paper, we present a novel eye location approach based on image context analysis. It is robust from the image variations such as illumination, glasses frame, and eyebrows. Image context of an image is any observable relevant attributes with other images. Image context analysis is carried out using the hybrid network of k-means and RBF. The proposed eye location employs context-driven adaptive Bayesian framework to relive the effect due to uneven face images. The appearance of eye patterns is represented by Haar wavelet. It also employs a merging and arbitration strategy in order to manage the variations in illumination and geometrical characteristics of ambient eye regions due to glasses frames, eye brows, and so on. The located eye candidates are merged or eliminated, and adaptive arbitration strategy is used based on a minimizing energy function by probabilistic forces and image forces. The adaptation is carried out by the analysis of image context. The experimental results show that the proposed approach can achieve superior performance using various data sets to previously proposed methods.

1 Introduction

Recently, automatic face recognition has become a hot topic of computer vision research area [2, 9]. And an important point in face recognition is face alignment. Face alignment includes spatial face normalize as scaling and rotating to match with face samples in the datasets. Already papers [10, 7] shown that face alignment has a huge effect on recognition accuracy. Contemporary, face alignment system usually perform eye location firstly. Most of the face recognition methods use eye positions which manually given. But for real time face recognition system, manually locating eye positions is obviously not reasonable. Therefore an automatic eye location method is needed for a real face recognition system. In this paper, we first propose a context-driven classifier. Our eye location method is then detection eyes using Bayesian discriminant method. And then we apply two kinds of post-processings to output of Bayesian classifier.

Under most cases, the eye center is measured with the pupil center. Contemporary eye location manners can be divided into two classes: active and passive eye location. The active location manners use restrictive representation of illumination. Under IR

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 540 – 549, 2006.

illumination, pupils display physical attributes which can be employed to localize eyes [2]. The benefits of active eye location manners are accurate and robust, but they need restricted lighting sources and they have many false locations with an outdoor condition, where the outdoor illumination influences the IR illumination. On the other hand, passive manners immediately detect eyes from images within normal illumination. The difficulties of passive manners in eye location are mainly caused by the variations of intrinsic object characteristics, viewpoint, viewing distance, eye brows, thick spectacle rims illumination.

In this paper we solve uneven illumination problem to employ context-driven framework and resolve viewpoint, eyebrows and spectacle rims problem with merging and arbitration strategy. The experiments have been carried out using the BioID, FERET and Inha face data sets. We achieved very encouraging results, especially for the face images with glasses and uneven illumination.

2 Image Context-Driven Eye Location

Our eye locator is designed to detect two eyes of 16×16 pixel minimum size, rotated up to ±20 degrees in image plane. It contains a pipeline of an image context-driven method, the Bayesian context aware method and the postprocessing strategy. This pipeline achieves automatically multi-resolution, feature representation and classification of eye candidates, in a single integrated scheme. The image contexts are affected by illumination direction, intensity, glasses frame, eyebrows, and so on. The image contexts are modeled and analyzed by the hybrid network of k-means and RBF. The image context analysis assigns detected face image into one of several image contexts so that Bayesian classifier and post-processing can be tuned in accordance with varying illumination and eye patterns. Hence, the proposed image context based eye location has advantages over previous approaches in robustness to varying environments.

2.1 The Proposed Approach

An overview of our eye location algorithm is described in Fig. 1, which contains three major modules: 1) image context analysis, 2) discriminant with context adaptive thresholds and 3) merging and arbitration strategy. The algorithm first classifies face images by class using k-means and RBF algorithm. The image which is classified by image context analysis passes through a Bayesian eye locator. The Bayesian eye locator consists of a six-level multi-resolution method, a feature vector generator using 2D Haar transform and an eye candidate windows classifier with the Haar feature based Bayesian discriminant method. There appear usually multiple eye candidate windows and they should be merged into two eye windows since there are two eyes in a face region. In this paper, the derived eye windows are converted into eye candidate centroids. The block diagram of the proposed method is shown in Fig. 1.

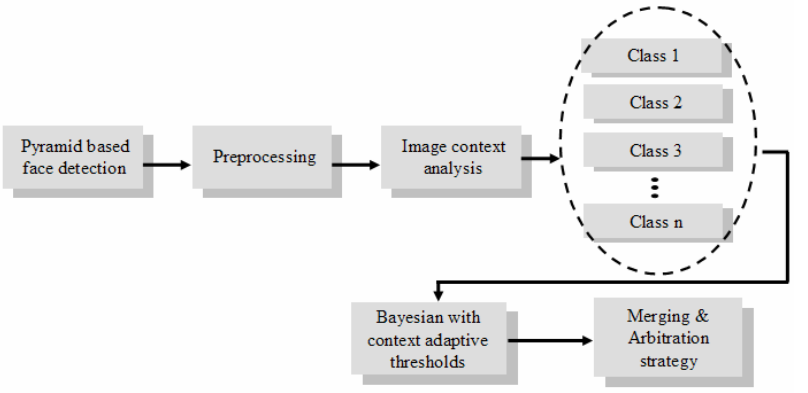


Fig. 1. Block diagram of the proposed eye location method

2.2 Image Context Analysis by the Hybrid Network of k-Means and RBF

Image context is any observable relevant attributes and its interaction with other images, and image context analysis deals with variations of illumination condition. We formulate the image context analysis to learn discriminative object patterns which can be measured formally, not intuitively as most previous approaches did. The image context analysis assigns detected face image into one of several image categories. The image contexts are modeled and analyzed by the hybrid network of k-means and RBF. The proposed image context-driven eye location has advantages over previous approaches in robustness to varying intrinsic and extrinsic illuminations.

The RBF networks, just like MLP networks, can therefore be used in classification and/or function approximation problems. In the case of a RBF network, we usually prefer the hybrid approach, described below [4]. The RBFs, which have a similar architecture to that of MLPs, however, achieve this goal using a different strategy. One cluster center is updated every time an input vector x is chosen by k-means from the input data set.

$$\begin{aligned}\hat{P}(\omega_i \mid x_k, \hat{\theta}) &= \frac{p(x_k \mid \omega_i, \hat{\theta})\hat{P}(\omega_i)}{\sum_{j=1}^{c} p(x_k \mid \omega_j, \hat{\theta}_j)\hat{P}(\omega_j)} \\ &= \frac{|\hat{\Sigma}_i|^{-1/2} \exp[-\frac{1}{2}(x_k - \hat{\mu}_i)^t \hat{\Sigma}_i^{-1}(x_k - \hat{\mu}_i)]\hat{P}(\omega_i)}{\sum_{j=1}^{c} |\hat{\Sigma}_j|^{-1/2} \exp[-\frac{1}{2}(x_k - \hat{\mu}_i)^t \hat{\Sigma}_i^{-1}(x_k - \hat{\mu}_i)]\hat{P}(\omega_i)}\end{aligned} \tag{1}$$

From Eq. 1, it is clear that the probability $\hat{P}(\omega_i \mid x_k, \hat{\theta})$ is large when the squared Mahalanobis distance is small. Suppose that we merely compute the squared Euclidean distance $\|x_k - \hat{\mu}_i\|^2$, find the mean $\hat{\mu}_m$ nearest to x_k and approximate $\hat{P}(\omega_i \mid x_k, \hat{\theta})$ as $\hat{P}(\omega_i \mid x_k, \hat{\theta}) \cong \begin{cases} 1 & \text{if } i = m \\ 0 & \text{otherwise.} \end{cases}$

Then the iterative application of $\hat{\mu}_i$ leads to the following procedure for finding $\mu_1, \mu_2, ..., \mu_c$. In the absence of other information, we may need to guess the "proper" number of clusters, *c*. Likewise, we may assign *c* based on the final application. Here and throughout, we denote the known number of patterns as *n*, and the desired number of cluster c. It is traditional to let c samples randomly chosen from the data set serve as initial cluster centers.

In training step we classify face images along illumination peculiarity, and in experience phase we apply Bayesian theory with optimized different thresholds at each cluster. We call this manner a context-driven method. The general Bayesian method [1] applies same threshold to whole test images, but the context-driven method applies several thresholds according to the number of clusters.

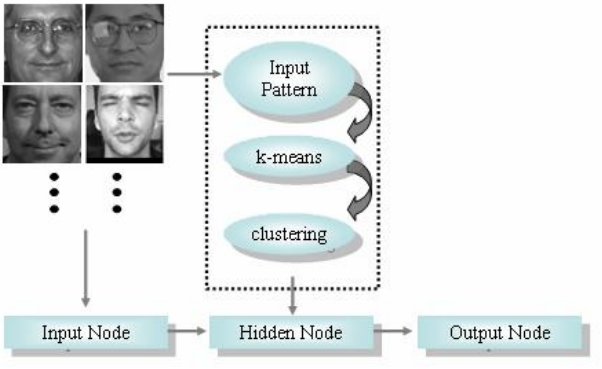


Fig. 2. Training system architecture

As showed Fig. 2, the idea is to train the network in two separate stages. In first stage, we perform an unsupervised training (k-means). In the second stage, the diverse thresholds are trained using the regular supervised approach. Input pattern is vectorized for grayscale image size of 16×16 pixels. The transformation from the input space to the hidden unit space is non-linear, whereas the transformation from the hidden-unit space to the output-space is linear. RBF classifier expands input vectors into a high dimensional space. RBF network has architecture that of the traditional three-layer back-propagation. In this paper, hidden units are trained using k-means. The network input consists of n normalized and rescaled size of 1/2 face images fed to the network as 1 dimension vector. And input unit has floating value [0, 1]. The vector value is normalized.

3 Haar Feature Based Bayesian Eye Candidate Location

In this paper, we accept 2D Haar wavelet representation for generating the feature vectors of eyes. Large number of eye images and non-eye images are needed to train the Bayesian classifier [1]. These images are normalized to the same scale (16×16). These samples are modeled using PCA for reducing feature dimension in the manner

of a paper [1]. Then, they are measured by Bayesian discriminant method with Mahalanobis distance [8]. At least two class models [1] are required to classify eye and non-eye using Bayesian method: one is eye class model and the other is non-eye class model. We employ the multiple two class models here for individual illumination contexts, respectively for each class.

3.1 Eye Candidate Window Using Haar Feature Based Bayesian Classifier

We employ Mahalanobis distance for each window. Let $d(x)_e$ be the Mahalanobis distance of the eye class, and $d(x)_n$ be that of for non-eye class. $d(x)_e$ and $d(x)_n$ can be calculated from the input pattern x, the eye class parameters (the mean eye, and the corresponding covariance matrix), and the non-eye class parameters (the mean non-eye, the corresponding covariance matrix) respectively. We use two thresholds, θ and τ as follows:

$$\begin{aligned} \theta &= \max(d(\alpha)_e) \\ \tau &= \max(d(\beta)_e - d(\beta)_n) \end{aligned} \tag{2}$$

where α is training samples of eye class and β is training samples of non-eye class. The two thresholds are constant values, which are calculated in the training time. The Bayesian classifier offers the classifying rule to the eye detection system as follows:

$$x \in \begin{cases} w_e & \text{if } d(x)_e < \theta) \text{ and } \left(d(x)_e - d(x)_n < \tau\right) \\ w_n & \text{otherwise} \end{cases} \tag{3}$$

4 Post-Processing for Deciding True Eye Centroid

The post-processing of the proposed eye location consists of the merging and arbitration strategies. Merging strategy was devised to improve the credibility of the eye location by removing false detections. Arbitration strategy finds a mediated position among multiple merged centroid patterns by a minimizing energy function of probabilistic forces and image forces.

4.1 The Proposed Merging and Arbitration Strategy

The result from the Bayesian classifier usually contains multiple eye candidate windows including false detections as illustrated in Fig. 3. Merging strategy is devised to improve the credibility of the eye location by removing false detections [4]. Extensive experiments discover that there are usually multiple candidate windows detected nearby eyes with high frequency, while false detections usually arise with low frequency. This discovery gives us a reasonable decision rule that can eliminate false detections and merge true detections into a single merged detection.

This method is good at not only improving accept rate but also at decreasing false detection. If a specific position is correctly classified as an eye, then all other detected positions which overlap it are regarded as errors, therefore these can be eliminated. The position with the higher number of detections is conserved, and the

position with the lower detections is eliminated. This method is controlled by two variances: the *threshold of overlapping density* and the *size of spread out*. We will only accept a detection if there are at least more than threshold detections within a region (spread out along x, y, and scale) in the detections phase. The *size* of the region is determined by a variable, which is the number of pixels from the center of the region to its edge. Each detected centroid at a particular position and scale is marked in image. Then, each position in image is replaced by the number of detections in a specified neighborhood of that position. This has the effect of spreading out the detections. The proposed merging strategy is shown in Fig. 4. In Fig. 4a ~ e, it shows simple case with overlap elimination process.

After the merging strategy is applied, sometimes two or more merged eye centroids appear. As a matter of fact, eyebrows or thick rims of spectacle often are so similar to closed eye that the proposed merging strategy sometimes makes a wrong decision. So both eye region and ambient eye regions should be considered together. The merging strategy enforces to remain a merged eye centroid which has a higher overlapping detections (i.e., higher density of overlapping virtual density) and to eliminate other merged eye centroid(s). However, if two or more high overlapping detections (merged centroids) independently survive around the eye, the merging strategy fails to produce correct eye location any more.

In order to resolve the above problem, we analyze the multiple merged window distribution and devise the proposed arbitration strategy. It is a method which finds a mediated position among false detections. The arbitration strategy is based on a minimizing energy function by probabilistic forces and image forces. The probabilistic forces serve to impose a reasonable mediated position according to their density of detections. The image force pushes the position toward salient eye features.

Representing the mediated position parametrically by $v=(x,y)$, we can use its energy function as:

$$E_{arbitration} = \int_0^1 E_{prob}(v) + E_{image}(v)ds \tag{4}$$

where E_{prob} represent the probabilistic energy of the detections due to density, E_{image} gives image force which consist of intensity and gradient. An example of the application of these functions is shown in Fig. 4. Fig. 4f shows fail of merging and overlap elimination strategy. Red marks are false merged eye windows which are made by merging strategy. A blue point of right region in Fig. 4g indicates that centroid points of each wrong eye windows. We generate one $n \times n$ rectangle region which has height and width which vertical distance between two centroid points as in Fig. 4h. The center of the region is became centroid of two centroid points. Energies of each pixel in the region are calculated and the pixel which has minimum energy is regarded as eye centroid. The energy functions introduction should be scaled so that each component of function contains comparable values.

This process is referred to as regularization. Each of the energy functions is adjusted to the range as follows:

$$e'_{xy} = \frac{e_{xy} - e_{\min}}{e_{\max} - e_{\min}} \tag{5}$$

where $e_{\min}$ and $e_{\max}$ are the minimum and maximum valued elements, respectively, in each energy function component.

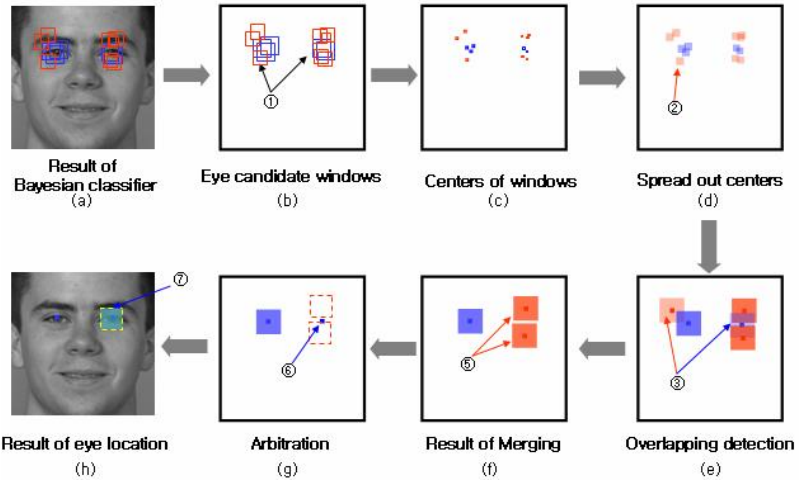


Fig. 3. The proposed post-processing (a) Result of Bayesian classifier. (b) Eye candidate windows. (c) Centers of eye candidate windows. (d) The windows are "spread out" and a merging is applied. If *threshold* is two, ② is conserved because its density is lower than *threshold*. (e) Eliminating overlapping detections. ③ is eliminated by elimination rule. (f) Result of merging. ⑤ will be arbitrated by arbitration strategy. (g) Arbitration. ⑥ is centroid of two false detections. (f) Result of proposed eye location. ⑦ is arbitration region.

5 Experiments

The training set is obtained from FERET database, and totally 1970 eyes of 985 faces are extracted and normalized for training. Experimental test set consists of BioID, FERET, and Inha data sets, and totally 6537 face images are used in our experiment.

There are several methods to measure the accuracy of eye location used in previous research [6, 3]. In this paper, we adopt a scale independent localization measurement, called relative accuracy of eye location, to estimate the accuracy of eye location [3]. The relative accuracy of detection is defined as follows [3]:

$$err = \frac{\max(d_l, d_r)}{d_{lr}} \tag{6}$$

In Fig. 4, we offer error curves of 3 test databases and the system can show satisfactory performance in the vicinity of err <0.14.

In order to make comparison with context-driven method with general Bayesian method, we cluster 1093 images FERET fa set for three, six and nine clusters.

Performance of three-clustering (k=3) is demonstrated in Table 2. It shows that context-driven method is superior to general Bayesian method. Table 3 is show similar results. From the tables, we can conclude that k-means based context-driven Bayesian method have an effect on eye location. Although it is a performance under *err* <0.10, Table 2 and 3 show similar performances with general Bayesian method of Table 1. We offer some examples out of the test sets for visual examination at Fig. 5. The system appears to be robust to the presence of glasses, closed eyes, slightly rotated faces and even significant pose changes. Fig. 5a shows that the proposed method is robust to the eye glasses and Fig. 5b also shows its robustness to illumination variation.

The proposed system is compared with other systems. We select paper [5] and [6] as targets of comparison because they use identical evaluation protocol (*err*) and data sets with us. In paper [6], the detection rate is 99.1% under err <0.20. On the other

Table 1. Performance of general Bayesian method with merging and arbitration strategy

Source	Images	Accepted faces	False detects	Acceptance rate
FERET	3816	3690	126	96.70%
BioID	1521	1458	63	95.86%
Inha	1200	1148	52	95.67%
Total	6537	6296	241	96.31%

err < 0.14

Table 2. Acceptance rate for three-clustering (k=3)

classes	images	general Bayesian method			context-driven method		
		detect	false	rate	detect	false	rate
class 0	293	277	16	94.54	281	12	95.90
class 1	672	650	22	96.73	650	22	96.73
class 2	128	119	9	92.97	124	4	96.88
total	1093	1046	47	95.70	1055	38	96.52

err < 0.10

Table 3. Acceptance rate for six-clustering (k=6)

classes	images	general Bayesian method			context-driven method		
		detect	false	rate	detect	false	rate
class 0	221	206	15	93.21	211	10	95.48
class 1	93	87	6	93.55	91	2	97.85
class 2	553	533	20	96.38	539	14	97.47
class 3	31	30	1	96.77	30	1	96.77
class 4	21	19	2	90.48	19	2	90.48
class 5	174	171	3	98.28	171	3	98.28
total	1093	1046	47	95.70	1061	32	97.07

err < 0.10

hand, our detection rate is 99.92% if err <0.20 at all test sets. In paper [5], the detection is considered to be correct if err <0.25. Their detection rate on BioID dataset is 94.81%. We evaluate proposed method on BioID under the same evaluation protocol. The detection rate of our system is 95.86% if err <0.14, and the detection rate is 98.82 % if err <0.25. And their system achieve on 97.18% of JAFFE data set. But backgrounds and illumination conditions of JAFFE are not as complex and diverse as these of BioID. Table 1 shows the comparative eye location performance of the proposed method in this paper and other methods.

Table 4. Comparative Eye Location Performance of the proposed method in this paper and other methods

Method	err < 0.2 (all dataset)	err < 0.25 (BioID)
Y. Ma	99.1%	N/A
H. Zhou	N/A	95.86%
The proposed method	99.92%	99.14%

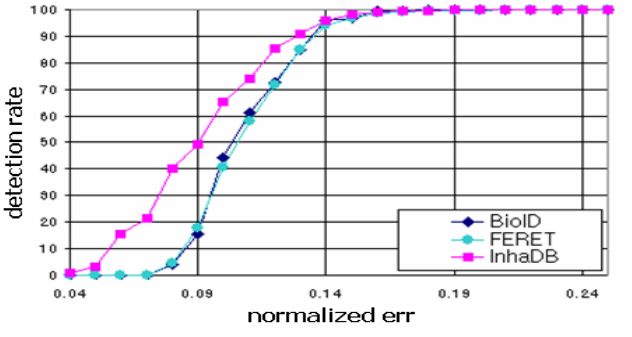
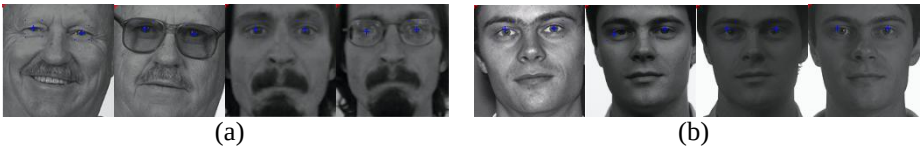


Fig. 4. Error curves



(a) (b)

Fig. 5. Eye location results on a subset of FERET, BioID photos

6 Conclusion

This paper addresses an image context-driven eye location method using the hybrid network of k-means and RBF. The proposed eye location employs image context based adaptive framework to relive the effect of uneven images coming from illumination condition, glasses frame, eyebrows, and so on. The novelty of this paper

comes from image context analysis based adaptive framework. First, an image context-driven method applied. Second, input images are classified by the context-driven Bayesian method. Third, we employ the merging and arbitration strategy. The image contexts are analyzed by c-means and RBF. In order to classify eyes and non-eyes, we use statistical modeling which evaluates the conditional probability density functions of eyes and non-eyes classes. Multiple Bayesian frameworks are constructed with different thresholds derived by image context. The adaptive merging and arbitration strategy is applied to improve accuracy of the system. Sophisticated post-processing, i.e. adaptive merging and arbitration strategy manages the variations in geometrical characteristics of ambient eye regions due to glasses frames, eye brows, and so on. The arbitration strategy is trained with 1970 eye images and 3844 random natural (non-eye) images. Experimental results using 6537 images (containing a total of 13074 eyes) from various image sources. The proposed method achieves 97.11 percent eye detection accuracy under err < 0.14.

References

[1] C. Liu, “A Bayesian Discriminating Features Method for Face Detection” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, no. 6, pp. 725-740, 2003

[2] A. Haro, M. Flickner, and I. Essa, “Detecting and tracking eyes by usingtheir physiological properties, dynamics, and appearance,” IEEEInternational Conference on Computer Vision and Pattern Recognition,vol. 1, 2000, pp. 163–168.

[3] O. Jesorsky, K. Kirchberg, R. Frischholz, "Robust face detection using the Hausdorff distance," In: J. Bigun, F. Smeraldi Eds. Lecture Notes in Computer Science 2091, Berlin: Springer, 2001, pp.90-95.

[4] M. H. Hassoun. “Fundamentals of Artificial Neural Networks,” MIT Press, 1995.

[5] H. Zhou, X. Geng, “Projection functions for eye detection,” Pattern Recognition, 2004, in press.

[6] Y. Ma, X. Ding, Z. Wang, N. Wang, "Robust precise eye location under probabilistic framework," IEEE International Conference on Automatic Face and Gesture Recognition, 2004

[7] A.M. Martinez, “Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class,” PAMI, IEEE Transactions on 24 (2002), no. 6, 748–763.

[8] Akihiko Watabe, Kazumi Komiya, Jun Usuki, Kayo Suzuki, Hiroaki Ikeda, "Effective Designation of Specific Shots on Video Service System Utilizing Mahalanobis Distance", IEEE Transactions on Consumer Electronics, Vol. 51, No. 1, February 2005

[9] Jeffrey Huang and Harry Wechsler, “Eye detection using optimal wavelet packets and radial basis functions (rbfs).,” International Journal of Pattern Recognition and Artificial Intelligence 13 (1999), no. 7, 1009–1026.

[10] P.J. Phillips, Hyeonjoon Moon, S.A. Rizvi, and P.J. Rauss, “The feret evaluation methodology for face-recognition algorithms,” IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (2000), no. 10, 1090–1104.

A Study on Vision-Based Robust Hand-Posture Recognition by Learning Similarity Between Hand-Posture and Structure

Hyoyoung Jang[1], Jin-Woo Jung[2], and Zeungnam Bien[3]

[1] Department of Electrical Engineering and Computer Science, KAIST,
373-1 Guesong-dong, Yuseong-gu, Daejeon, 305-701, Korea
hyjang@gmail.com
[2] Human-Friendly Welfare Robot System Research Center, KAIST,
373-1 Guesong-dong, Yuseong-gu, Daejeon, 305-701, Korea
jinwoo@ctrsys.kaist.ac.kr
[3] Department of Electrical Engineering and Computer Science, KAIST,
373-1 Guesong-dong, Yuseong-gu, Daejeon, 305-701, Korea
zbein@ee.kaist.ac.kr

Abstract. This paper proposes a robust hand-posture recognition method by learning similarity between hand-posture and structure for the performance improvement of vision-based hand-posture recognition. The difficulties in vision-based hand-posture recognition lie in viewing direction dependency and self-occlusion problem due to the high degree-of-freedom of human hand. General approaches to deal with these problems include multiple camera approach and methods of limiting the relative angle between cameras and the user's hand. In the case of using multiple cameras, however, fusion techniques to induce the final decision should be considered. Limiting the angle of user's hand restricts the user's freedom. The proposed method combines angular features and appearance features to describe hand-postures by a two-layered data structure and includes learning the similarity between the two types of features. The validity of the proposed method is evaluated by applying it to the hand-posture recognition system using three cameras.

1 Introduction

Research on automatic hand-posture recognition is important to realize natural human-computer interaction. Most people use hand-postures in addition to words when they speak: clenching a fist, pointing, pinching, and doing finger spelling. Thus hand-posture recognition is attracting attention as one of the intuitive human-computer interfaces. This paper presents a robust method to recognize hand-postures captured by vision sensors.

In vision-based tracking, visibility is important because many occlusions occur. Cameras should be placed so that hands being tracked remain in the scene. The number of cameras is another important issue. Although using multiple cameras increase the algorithmic complexity of dealing with multiple image streams, they

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 550 – 559, 2006.

improve visibility. Once raw data have been collected, the system analyzes the existence of hands in the scene and recognizes hand-postures. A recognition technique should be chosen on the basis of how many postures are in the recognition set, the complexity of the set, and whether or not the set is known beforehand.

The work on hand-posture recognition is divided into two streams of research: model-based and view-based approaches.

Model-based approaches use an articulated three-dimensional (3D) hand model to describe hand-postures[1][2]. The model is projected onto the image and an error function is computed, scoring the quality of the match. The model parameters are then adapted such that the error is minimized. Mostly, it is assumed that the model configuration at the previous frame is known, and only a small parameter update is necessary. Therefore, model initialization is important in model-based approaches. In many researches, the model is aligned manually in the first frame.

In the view-based approach, the problem is formulated as a classification problem[3][4]. A set of hand feature is labeled with a particular hand posture, and a classifier is trained with this dataset. These techniques have been employed for gesture recognition tasks, where the number of learnt postures is relatively limited. To recognize a limited set of hand-postures, only useful models are registered. Because calculation for the 3D features is not necessary, it takes less computation. However, it requires images from various viewing directions to define hand-postures. If a hand-posture is obtained from a view that is not predefined, this method cannot guarantee good performance.

This paper proposes the Object Shape and Structure Network (OSS-Net) to recognize hand-postures under varying viewpoints. It is a layered hand-posture recognition structure in which view-based and model-based approaches are complementarily combined. In this method, there are two advantages. First, it considers both changes of the shape due to viewpoint variation and the angular configuration of the hand. Since the relation between two-dimensional (2D) appearance features and three-dimensional (3D) structures is many to many, they should be considered together to improve the robustness against viewing direction changes. Second, it gives a solution to deal with multiple image streams.

Fig. 1 shows the target hand-postures.

Fig. 1. Target hand-postures

To recognize the hand-postures in Fig. 1, three cameras equipped on the ceiling were used (Fig. 2). Fig. 3 shows the effect of variance in viewpoint. Although their appearances are different from each other, those five images have been taken for the same hand-posture.

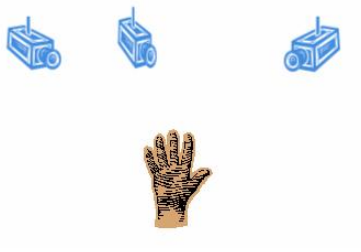


Fig. 2. Camera configurations

Fig. 3. Example of hand appearances from various

The remainder of the study is organized as follows: Section 2 shows the proposed hand-posture database structure, Section 3 deals with details of the hand-posture recognition system application, Section 4 shows some experimental results and Section 5 concludes.

2 Hand-Posture Database Structure

We propose *Object Shape and Structure Network (OSS-Net)* model to recognize hand-postures under varying viewpoints.

OSS-Net model is a hand-posture database structure that has both characteristics of the 3-dimensional model based approach and appearance-based approach. It is a layered hand-posture recognition structure in which view-based and model-based approaches are complementarily combined. In this method, there are two advantages. First, it considers both changes of the shape due to viewpoint variation and the angular configuration of the hand. Since the relation between two-dimensional (2D) appearance features and three-dimensional (3D) structures is many to many, they should be considered together to improve the robustness against viewing direction changes. Second, it gives a solution to deal with multiple image streams.

The OSS-Net is defined as below:

Def. 1 Object Shape and Structure Network (OSS-Net),

OSS-Net = (*S*, *P*, *L1*, *L2*, *C*)

S: pattern set

P: network variable set

L1: 2D view-based feature layer

L2: 3D model-based feature layer

C: set of interconnections between the elements of *L1* and *L2*

The main parts of the structure are 2D view-based feature layer (*L1*), 3D model-based feature layer (*L2*), and the set of interconnections between the elements of *L1* and *L2* (*C*). Each layer is constructed by several nodes representing a certain type of hand-posture. Nodes composing layers are described by a feature vector and a connectivity vector which defines connectivity to hand-posture model identifiers. A hand-posture is described by interconnections between the nodes. By constructing nodes on *L1* layer with appearance features observed from various views, we can consider the varying view effects on the shape. The connections between nodes on *L2* layer represent different hand-postures with the same appearance. Pattern set S is the set of data used to construct OSS-Net. Network variable P defines network parameters such as threshold values and maximum number of repetition.

L1 and *L2* layers are defined as below:

Def. 2 2D View-based Feature Layer, $L1 = (X, U)$

$X = \{\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_m\}$: 2D view-based feature node set

m: the number of view-based feature nodes

$\underline{x}_i = (\underline{xf}_i : \underline{xh}_i)$, $i = 1, \cdots, m$,

$\underline{xf}_i$: 2D view-based feature vector

$\underline{xh}_i$: p-dimensional vector representing connectivity to the hand-posture where p is the number of hand-postures to be recognized

$U = \{u_{ij}\}$, $i, j = 1, \cdots, m$: Interconnections between view-based feature nodes

m: the number of view-based feature nodes

$$u_{ij} = \begin{cases} 0 & if \quad i = j \\ p_u(i, j) & otherwise \end{cases},$$

$p_u(i, j)$: connectivity between $\underline{x}_i$ and $\underline{x}_j$

$\underline{xh}_i$ represents the relation between i[th] appearance-based representation and each hand-postures. There can be several possible hand-postures which results in the same appearance. However, for example, if the value of j[th] element in $\underline{xh}_i$ is bigger than others, the j[th] element is considered the most probable hand-posture for i[th] appearance in *L1* layer.

Def. 3 3D Model-based Feature Layer, $L2 = (Y, V)$

$Y = \{\underline{y}_1, \underline{y}_2, \cdots, \underline{y}_n\}$: 3D-model based feature node set

n: the number of 3D-based feature nodes

$\underline{y}_i = (\underline{yf}_i : \underline{yh}_i)$, $i = 1, \cdots, n$

$\underline{yf}_i$: 3D-model based feature vector

$\underline{yh}_i$: p-dimensional vector representing connectivity to the hand-posture where p is the number of hand-postures to be recognized

$V = \{v_{ij}\},\ i, j = 1, \cdots, n$: interconnection between 3D model-based feature nodes

n : the number of 3D model-based feature nodes

$$v_{ij} = \begin{cases} 0 & if \quad i = j \\ p_v(i, j) & otherwise \end{cases},$$

$p_v(i, j)$: connectivity between $\underline{y}_i$ and $\underline{y}_j$

The meaning of each connection between nodes is similar with the case of *L1* layer. $\underline{yh}_i$ is the relation between a 3D model-based feature nodes and each hand-posture model.

The interconnection between layers is defined as follows:

Def. 4 Connectivity between Layers, C :

$$C = \{c_{ij}\},\ i = 1 \cdots n,\ j = 1, \cdots m$$

$$c_{ij} = \begin{cases} 1 & if \quad x_j \quad is \quad related \quad to \quad y_i \\ 0 & otherwise \end{cases}$$

The elements of *C* indicate the interconnection strength between view-based nodes in *L1* layer and 3D model-based nodes in *L2* layer.

With the hand-posture database based on OSS-Net model, the hand-posture recognition process copes with search within the structure.

To construct OSS-Net, two kinds of hand-posture descriptors, view-based and 3D model-based, are required to construct each of two layers which comprise the network. For view-based feature layer, we have used the boundary descriptor. It is a real vector of which elements are lengths between the center of segmented hand region and its boundary pixels. For 3D model-based feature layer, finger-angles are measured by a dataglove. The dataglove is only used in the training phase. Once trained, the system works using the 3D-model based feature layer with parameters decided from training.

3 Hand-Posture Representation and Recognition

The OSS-Net explained in the previous chapter was applied to build the hand-posture recognition system. Hand-posture descriptions from two aspects are needed to construct two layers which comprise the database. We used boundary description feature vector composed of lengths between the center of the hand region and the boundary pixel and finger-angles measured by a dataglove. Fig. 4 shows boundary description feature extracting process.

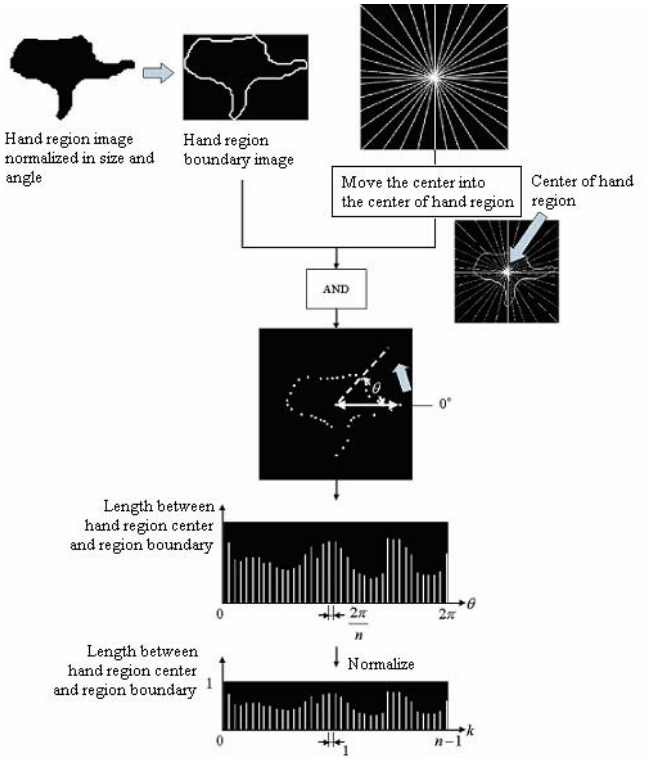


Fig. 4. Boundary description extracting process

First, the segmented hand-region is normalized in size and aligned to the principle axis. Next, the region boundary is achieved by the conventional boundary following algorithm. We use an angle segmentation mask to select points on boundary with same interval. After the center of an angle segmentation mask is moved onto the hand region center, logical *AND* operation is applied. From the resulting image with dots which mean same interval point on boundary, we can calculate the length between those dots and region center. It is expressed as *n*-dimensional vector where *n* is the number of angle segmentation. Finally it is normalized to have 1 as the maximum value.

Fig. 5 shows a dataglove used in finger-angle extraction process.

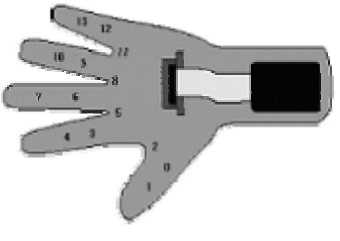


Fig. 5. Dataglove (0~13: Sensor position to achieve angles)

The finger-angle values are normalized by equation (2).

$$normalized\_data = \frac{val - \min}{\max - \min} \times \max \tag{2}$$

val: input value for the sensor
max: the maximum value of the sensor input
min: the maximum value of the sensor input

The boundary descriptors and normalized finger-angle vectors are used to construct the sets, $\{\underline{xf}_i\}$ and $\{\underline{yf}_j\}$, $i=1,\ldots,n$, $j=1,\ldots,m$, respectively. In the construction stage, the hand-posture identifiers are given together with these two kinds of feature vectors.

First, we find a node on *L1* with a view-based feature vector which has the smallest Euclidean distance from the input view-based feature vector $\underline{xf}_i$. If the distance value is bigger than the predefined threshold value, create a new node on *L1*. Otherwise, the selected node is adjusted to have a feature vector which is the center of three vectors: the most similar feature vector, the next most similar feature vector, and the input feature vector. If a new node is created on *L1*, the same procedure is conducted on *L2* for the corresponding input finger-angle vector.

After creating or adjusting a node, connections between nodes are modified. There are two kinds of connections: *intralayer* and *interlayer* connections. *U* in Def. 2 and *V* in Def. 3 are intralayer connections. *C* in Def. 1, interconnection between the elements each from *L1* and *L2* layer, is the interlayer connection. Intralayer connections reflect the possibility that they are caused by the same hand-posture. Intralayer connections reflect two cases by real values. One is due to the variance in the viewing direction. Even the same appearance may be caused by 3D structures different from each other, that is, a different hand-posture. Likewise, the same 3D structure does not always show the same appearance. Interlayer connections reflect these two cases.

Under the OSS-Net model, hand-posture recognition is processed by transition between nodes constructed by given data. Q-learning algorithm is applied to determine the transition. The comparison is occurred on the *L1* layer and then, transition between nodes on the *L2* layer is conducted according to the previous result. The problem description by Q-learning is shown below:

- Initial values of Q-table: Immediate reward values
- Action: Transition between nodes
- State: Allocated value for each feature node determined by Euclidean distance between nodes and transition
- Immediate Reward:

  -5 for false and normalized value of inverse Euclidean distance for true

Fig. 6 shows hand-posture recognition process based on OSS-Net database structure using three cameras.

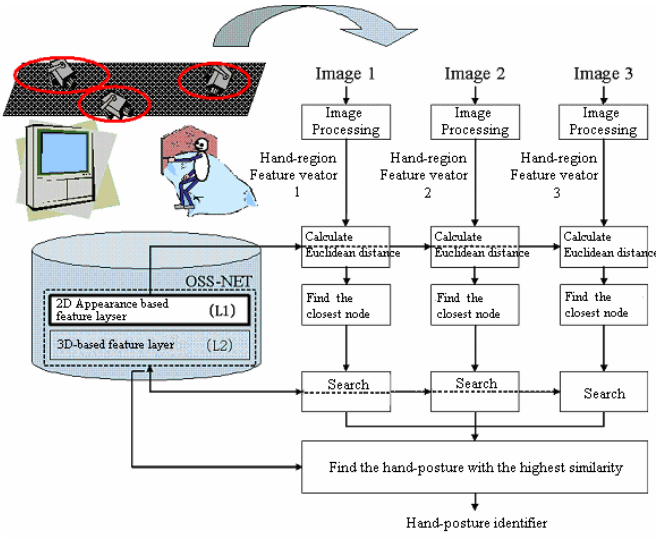


Fig. 6. Hand-posture recognition process

4 Experimental Results

The hand-posture recognition system has been constructed and tested using three cameras. To build the hand-posture database, subjects were requested to pose 15 orientations for each hand-posture as Fig. 7. The sample was acquired from 3 people and 5 trials for each. For each trial we can get 225 images: three cameras, 5 hand-postures and 15 views. Consequently we get 3375 images. During data capture stage, a dataglove is used to get the finger-angle data. For recognition, a dataglove is not used. Randomly selected 1500 images (300 images/hand-posture) were used to build OSS-Net structure and randomly chosen 1500 images (300 images/hand-posture) from remaining were used to test.

From experiments, we decided 0.55 as the exploration probability for Q-learning. Generally, higher exploration probability results in better recognition rate and higher

Fig. 7. Fifteen orientations for each hand

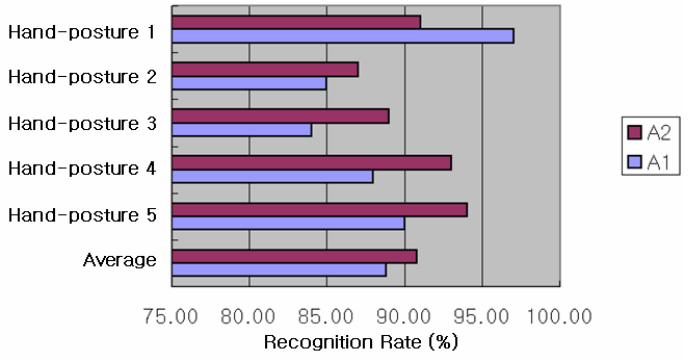


Fig. 8. Hand-posture recognition result (A1: Similarity by Euclidean distance, A2: Similarity by Q-learning)

convergence speed. However, too high exploration probability makes recognition rate rapidly decrease and increases the number of iteration for convergence.

To validate the proposed similarity defining method using Q-learning, we also tested the case using Euclidean distance as a similarity measure. As Fig. 8 shows, the proposed method has showed good performance except for the hand-posture 1(fist). For the constructed OSS-Net, we tested 10 times and got average recognition rate of 90.80%. The recognition failure of the hand-posture 1 was caused by commonness of the appearance. Since, according to the viewing direction, other 4 hand-postures also can be seen as the 'fist-like' shape, 5 hand-posture sets commonly include the appearance similar to a fist.

5 Conclusion

A human-hand is the object with the high articulation, and thus, it has high degree of freedom. This, together with self-occlusion phenomena and view-variance, makes it hard to recognize a human-hand. This paper suggested OSS-Net structure, two-layered database structure representing both 2-dimensional appearance features and 3-dimensional features. The contribution of this paper is a robust hand-posture recognition method with characteristic feature database structure, OSS-Net, and a novel method to combine multiple views.

Although the proposed method gives effective way of hand-posture recognition, more research is needed for optimization of each stage. Future work will focus on the numerical extension of the database and advanced analysis on features and distance measure.

Acknowledgement

This research is supported by the Ministry of Commerce, Industry and Energy and the Human-friendly Welfare Robot System Engineering Research Center of KAIST (R11-1999-008).

References

1. J.M. Rehg, T. Kanade: Visual Tracking of High DOF Articulated Structures - an Application to Human Hand Tracking. Proc. of ECCV'94 (1994) 35-46
2. D. Lowe, Fitting Parameterized: Three Dimensional Models to Images. IEEE Trans., PAMI, Vol.13, No.5 (1991) 441-450
3. B. Moghaddam, A. Pentland: Maximum Likelihood Detection of Faces and Hands. Proc. of Int. Workshop on Automatic Face and Gesture Recogtnition (1995) 122-128
4. U. Brockl-Fox: Realtime 3-D Interaction with up to 16 Degrees of Freedom from Monocular Video Image Flows, Proc. of Int. Workshop on Automatic Face and Gesture Recognition (1995) 172-178

Kernel-Based Method for Automated Walking Patterns Recognition Using Kinematics Data

Jianning Wu, Jue Wang, and Li Liu

Key Laboratory of Biomedical Information Engineering of Education Ministry,
Xi'an Jiaotong University, Xi'an, 710049 China
jianningwu@mail.xjtu.edu.cn

Abstract. A novel scheme is proposed for training Support Vector Machines (SVMs) in automatic recognition of young-old gait types with a higher accuracy. Kernel-based Principal Component Analysis (KPCA) is employed to initiate the training set, which efficiently extracts more nonlinear features from highly correlated time-dependent gait variables and improves the generalization performance of SVM. With the proposed method (abbreviated K-SVM), the gait patterns of 24 young and 24 elderly normal participants were analyzed. Cross-validation test results show that the generalization performance of K-SVM was on average 89.6% to identify young and elderly gait patterns, compared with that of PCA-based SVM 83.3%, SVM 81.3% and a neural network 75.0%. These results suggest that K-SVM can be applied as an efficient gait classifier for young and elderly gait patterns.

1 Introduction

It is well known that aging influences gait patterns, and considerable studies have reported changes during obstructed and unobstructed walking, which indicate age-related declines in lower limb control [1]. In [2], Nigg et al. studied the effects of aging through analyzing joint angular motion data such as ankle, knee and hip joint angles. In the clinical context, identification of the significant changes in gait between young and elderly is necessary to prevent severe injuries such as falls and alzheimer disease [3]. Thus, automated recognition of gait change has many advantages such as early identification of at-risk gait and monitoring the progress of treatment outcomes [4].

Recently, the machine classifier becomes more and more prevailing on the application of classification /recognition, which provides an opportunity to automated recognition of gait pattern changes with a higher accuracy. The typical Neural Network (NN) has been adopted to classify different kinds of gait patterns. For example, in the paper [5], Lees and Barton employ NN to discriminate simulated gait (e.g., leg length discrepancy) using features from lower-limb joint-angle measures. Holzreiter and Kohle [6] applied NNs for classification of normal and pathological gait using force platform recordings of foot-ground reaction forces. Especially, Rezaul K. Begg introduced Support Vector Machine (SVMs) to classify the gait

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 560 – 569, 2006.

patterns and gained a superior performance of classification. Although these above methods can classify the gait data with a relative higher accuracy, they all ignore the heuristic nature of these machines for which proper pre-processing of input variables are essential for good generalization [7].

In this study, we proposed a novel scheme of training SVM for automated recognition of young-old gait patterns using joint angular motion such as the ankle, knee and hip joint angles. As we know, the characteristics of gait data, such as high-dimensionality and nonlinear relationship, make the automated recognition by a machine classifier a challenging endeavor [7]. Pre-processing the raw gait data for machine classifier is critical for the entire result of gait analysis. Considering the high-dimensionality and nonlinear nature of gait data, we adopted kernel-based principal component analysis (PCA) to extract nonlinear features from highly correlated time-dependent gait variables, and then to train the SVM. The advantage of using kernel method is that it is neither necessary to know the form of the function, nor to calculate the dot product in the higher-dimensional space due to fact that an inner product in the feature space has an equivalent kernel in the input space. Thus, the kernel-based gait model was produced to classify different kinds of gait patterns.

This paper is organized as follows: Section 2 presents the procedure of the application of KPCA to process gait data. In Section 3, we briefly explain why we choose SVM as a machine classifier for gait data. In Section 4, we evaluate the performance of our technique via experiments. Discussions and conclusions are given in the Section 5.

2 KPCA for Gait Data Pre-processing

KPCA [8], proposed by Scholkopf et al, is a technique of generalizing linear PCA into nonlinear case by using the kernel method. Its basic idea is that linear PCA is performed in some high-dimensional feature space which is related to the input space by a possibly nonlinear map, just as a PCA in input space. Here, we adopt KPCA to extract nonlinear features from gait variables as follows.

Given a set of M centered gait data x_k, $k=1,...,M$, $\mathrm{x}_k \in R^N$, $\sum_{k=1}^{M} \mathrm{x}_k = 0$.

KPCA is to first map each gait data x_k into the higher-dimensional feature space F than the original one via a nonlinear function ϕ

$$\phi : R^N \rightarrow F \tag{1}$$

Then, the PCA problem in F can be formulated as the diagonalization of an M-sample estimate of the covariance matrix

$$\hat{\mathrm{C}} = \frac{1}{M} \sum_{i=1}^{M} \phi(\mathrm{x}_i) \phi(\mathrm{x}_i)^T \tag{2}$$

where $\phi(\mathrm{x}_i)$ are centered nonlinear mapping of the input variables $\mathrm{x}_k \in R^N$, $\sum_{k=1}^{M} \mathrm{x}_k = 0$. Here, we have to find eigenvalues $\lambda \geq 0$ and non-zero eigenvectors V, satisfying the equation

$$\lambda \mathrm{V} = \hat{\mathrm{C}}\mathrm{V} \tag{3}$$

Note that all solutions V with $\lambda \neq 0$ lie in the span of mappings $\phi(\mathrm{x}_1),...,\phi(\mathrm{x}_M)$. Consequently, the equivalent relation can be written as

$$\lambda\left(\phi(\mathrm{x}_k)\cdot \mathrm{V}\right) = \left(\phi(\mathrm{x}_k)\cdot \hat{\mathrm{C}}\mathrm{V}\right) \quad \text{for all } k = 1,...,M \tag{4}$$

Also, there are coefficients α_i $(i = 1,...,M)$ such that

$$\mathrm{V} = \sum_{i=1}^{M} \alpha_i \phi(\mathrm{x}_i) \tag{5}$$

Combination of Eq.(2),(4) and (5) yields

$$\lambda \sum_{i=1}^{M} \alpha_i \left(\phi(\mathrm{x}_k)\cdot\phi(\mathrm{x}_i)\right) = \frac{1}{M}\sum_{i=1}^{M} \alpha_i \left(\phi(\mathrm{x}_k)\cdot\sum_{j=1}^{M}\phi(\mathrm{x}_j)\right)\left(\phi(\mathrm{x}_j)\cdot\phi(\mathrm{x}_i)\right) \tag{6}$$
$$\forall k = 1,...,M$$

Further, we define an $M \times M$ kernel matrix K such that

$$K_{ij} := \left(\phi(\mathrm{x}_i)\cdot\phi(\mathrm{x}_j)\right) = K(\mathrm{x}_i, \mathrm{x}_j) \tag{7}$$

Here, kernel function $K(\mathrm{x}_i, \mathrm{x}_j)$ is introduced so that the mapping of $\phi(\mathrm{x}_i)$ from x_i is implicit. As K is symmetric, it has a set of Eigenvectors which span the complete space, thus

$$M\lambda\alpha = \mathrm{K}\alpha \tag{8}$$

Therefore, we only need to diagonalize K to obtain the normalization condition for $\alpha^p,...,\alpha^M$

$$\lambda_k\left(\alpha^k\cdot\alpha^k\right) = 1 \tag{9}$$

Finally, we can extract principal components by computing the projection of $\phi(\mathrm{x})$ onto the eigenvector V^k in high-dimensional space $F\left(k = p,...,M\right)$.

$$\left(\mathrm{V}^k \cdot \phi(\mathrm{x})\right) = \sum_{i=1}^{M} \alpha_i^k K\left(\mathrm{x}_i, \mathrm{x}\right) \tag{10}$$

Therefore, we only choose the first n nonlinear principal components, e.g. the directions which describe a desired percentage of data variance, and thus work in an n-dimensional sub-space of feature space F. This allows us to construct a novel gait classifier, K-SVM called in this study, where a preprocessing layer extracts nonlinear gait features for classification of gait patterns later.

3 SVM Classifier for Gait Patterns Recognition

Support Vector Machine, proposed by Vanpick and based on the VC theory and SRM, is one prevailing tool for machine learning. The SVM implements the following idea: it maps the input data nonlinearly into a higher-dimensional feature space where an optimal separable hyper-plane is constructed via kernel function. The more detailed description can be found in [9].

In this study, considering the situation of statistically small size issue due to the fact that gait data collection is costly, we adopted the SVM for it is non-sensitive to the sample size and the dimensions. Thus, we first extract non-linear principal components according to Eq. (10), and then train a Support Vector Machine. Combining principal components extracted with the Support Vector decision function, we thus obtain machines of the type

$$f\left(\mathrm{x}\right) = \mathrm{sgn}\left(\sum_{i=1}^{M} \beta_i K_2\left(g\left(\mathrm{x}_i\right) \cdot g\left(\mathrm{x}\right)\right) + b\right) \tag{11}$$

with

$$g\left(\mathrm{x}\right) = \left(\mathrm{V}_j \cdot \phi(\mathrm{x})\right) = \sum_{k=1}^{M} \alpha_k^j K_1\left(\mathrm{x}_k \cdot \mathrm{x}\right) \tag{12}$$

where $K\left(\mathrm{x},\mathrm{y}\right)$ is a positive definite symmetric function, b is a bias estimated on the training set. Note that different kernel function K_1 and K_2 are used for KPCA and SVM respectively. Here, we can obtain the expansion coefficients β_i which are computed by a standard Support Vector Machine. The more detailed solution for β_i is given in [9].

4 Experiments and Results

4.1 Acquisition and Features Selection of Gait Data

The experiment was carried out within two differently aged groups of subjects, one of which included twenty-four healthy young men with mean age (27.5±5.3) years and

heights (172±5.4) cm, while the other included twenty-four healthy elderly ones with mean age (62.4±4.3) years and heights (169±4.8) cm. The subjects had no known injuries or abnormalities affecting their gait. The gait was recorded using 3D-Optotrak® (NDI Corp. Canada). The motion tracking system was composed of Optotrak3020 motion sensor, Optotrak3020 control unit, Optotrak camera and markers, and IBM computer. With markers attached to lower limb joints and segments [10], motion analysis system collected the gait data. Gait recordings were performed during comfortable walking on a 10m laboratory walkway, and the sample frequency was 100Hz. All subjects in this study completed 3 gait trials and the kinematic parameters of each subject were calculated using the mean of 3 trials, and temporal data were normalized to gait cycle. A subject's sagittal angles plots during gait cycle is illustrated in Fig.1 (0%~60% stance phase, 60%~100% swing phase).

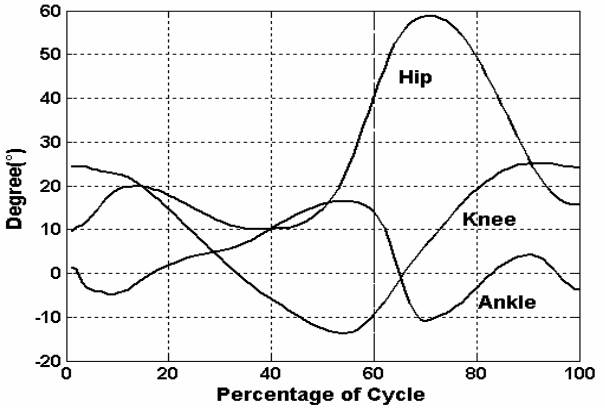


Fig. 1. The sagittal angles of the hip, knee and ankle joint of the subject

According to the definition of gait feature parameters by Sandro Giannini and clinical situation [11], we selected the gait feature parameters based on the joint-angle curve within a gait cycle. (1) joint angle at heel contact; (2) joint angle at toe off; (3) joint angle at flat foot;(4) joint angle at heel off;(5) maximal flexion angle during swing phase;(6) maximal extension angle during stance phase;(7) joint angular range of motion during the stance phase; (8) joint angle joint angular range of motion during the swing phase. Thus, all 24 features described the gait characteristics.

4.2 KPCA for Gait Data Analysis

The gait nonlinear feature was extracted from gait data by using KPCA algorithm. In this section, we compared the performance of KPCA with different kernel functions such as Gaussian RBF and the Polynomial function. Forty-eight subject's gait data with all 24 features were adopted as a sample set, and randomly selected 40 (20 young and 20 elderly) and 8 (4 young and 4 elderly) subjects' data were used as a training and testing sample set, respectively. According to section 2, its algorithm step is as follow:

(1) Compute the dot product matrix K_1 according to Eq. (7).

(2) Compute the Eigenvector expansion coefficients α_i by requiring Eq. (8) and (9).

(3) To extract the principal components of a test gait data x, we then compute projection onto the Eigenvectors by Eq. (10).

As a result, the nonlinear information feature of gait can be obtained easily by a simple transformation for the projected data. We evaluate the ability of KPCA to preprocess the raw gait data as follows. First, the parameters of SVM were set at the fixed values. Then, we observed the accuracy of classification under the condition of the various extraction features using KPCA and PCA respectively. And also we compared the results to SVM without using KPCA or PCA.

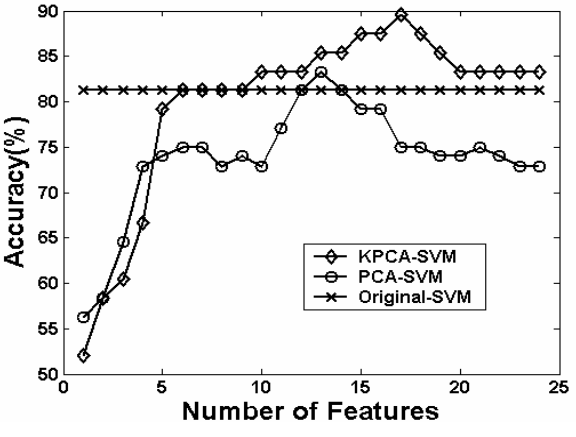


Fig. 2. The Relation between Accuracy-No. of Feature

Fig.2 illustrates the extent, to which change in the selection of the number of feature dimension can influence classification accuracy. The accuracy varies with the different number of the feature dimensions. When the dimensionality is more than 5, KPCA-based SVM shows a better performance than that of PCA-based SVM. The maximal classification accuracy (89.6%) is reached while the number of feature dimensions equals to 16, whereas the maximal classification accuracy of PCA-based SVM is 83.3% while the number of feature dimensions equals to 14.

4.3 Evaluation of K-SVM

After extracting non-linear principal components, we trained and tested a SVM classifier. In the section, we evaluate the performance of KPCA-based SVM, which we abbreviatedly call K-SVM. Training SVM classifier mainly includes initialization of the training set or samples, and optimization of parameter such as regularization parameter C, the Gaussian parameter δ and number of principle components n of KPCA and the kernel parameter q of SVM (SVM with RBF, q denotes the width of

RBF function; SVM with Poly, q denotes the degree of polynomial). We describe detailed training procedure as follow:

(1) Initiate the parameters C, δ, n and q, and construct an initial training set through KPCA pre-processing technique.
(2) The cross validation method is used to select the optimal free parameters C, δ, n and q till an optimal SVM Gait Classifier are produced.

The cross-validation is adopted to adjust the above parameters. The Evaluation Criteria is described in section 4.3.2. And 4.3.3 presents comprehensive comparative results of K-SVM and other types of machine classifiers.

4.3.1 Cross-Validation

Cross-validation is a standard test commonly used to test the ability of the classification system using various combinations of the testing and training data sets [12]. As the number of gait data available was limited in this experiment, a 6-fold cross-validation test was applied, in which 48 subjects' data were divided into six segments with the testing data set (8) selected as: Segment 1 (1-8), Segment 2 (9-16), Segment 3 (17–24), Segment 4(25-32), Segment 5(33-40) and Segment 6(41-48). Each of the six cross-validation test segments, therefore, had 4 young and 4 elderly subjects while their respective training segment included the remaining 20 young and 20 elderly subjects' data.

4.3.2 Evaluation Criteria

The following three measures of accuracy (Ac), sensitivity (Se) and specificity (Sp) were used to assess the performance of the SVM classifier [13].

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \tag{13}$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \tag{14}$$

$$Specificity = \frac{TN}{TN + FP} \times 100\% \tag{15}$$

where TP is the number of true positives, i.e. the SVM recognizes an elderly gait labeled as elderly; TN is the number of true negatives, i.e., SVM identifies a young gait labeled as young. FP is false elderly identifications; FN is false young identifications. Accuracy indicates overall detection accuracy; sensitivity is defined as the ability of the classifier to recognize an elderly gait pattern whereas specificity would indicate the classifier's ability not to generate a false detection (normal young gait).

4.3.3 Comparative Results

In Table 1, we compare the performance of K-SVM to the other types of SVM classifiers, including original SVM and PCA-SVM. In any kind of classifier, three types of kernel functions were adopt, including Poly, RBF and Linear. The parameters such as regularization parameter C, the kernel parameter of KPCA δ and number of principle components n of KPCA and the kernel parameter q of SVM, were all set the optimal values. According to the records, the K-SVM (Poly kernel for KPCA; Linear kernel for SVM) gained the best performance with the Ac of 0.89, Sp of 0.89, and Se of 0.89, followed by the PCA-SVM and then the original SVM. We also conclude that the type of kernel function of SVM is still non-sensitive to the performance of SVM classifier. Also, the Neural Network with three-layer was selected as an evaluation purpose of K-SVM. The NN adopt backpropagation (BP) as its learning algorithm. We observed that the Ac, Sp and Se of K-SVM are great more than those of NN. This demonstrates K-SVM is a robust classifier for gait data.

Table 1. The comparative classification performance of SVM, K-SVM and PCA-SVM

Techniques	Kernel function of SVM	(C, δ, q)	(Ac, Sp, Se)	n
SVM	Poly	(0.25,450,3)	(0.81,0.81,0.79)	---
	Linear	(0.3,400,---)	(0.81,0.81,0.81)	---
	RBF	(0.5,500,630)	(0.83,0.81,0.81)	---
K-SVM	Poly	(0.6,600,3)	(0.87,0.87,0.85)	13
	Linear	(0.7,700,---)	(0.89,0.89,0.89)	16
	RBF	(0.65,800,950)	(0.87,0.89,0.87)	15
PCA-SVM	Poly	(0.75,650,3)	(0.83,0.81,0.83)	11
	Linear	(0.55,600,---)	(0.83,0.83,0.83)	14
	RBF	(0.8,850,900)	(0.83,0.83,0.81)	12
NN (BP)	----	----	(0.75,0.75,0.75)	----

Note: C, q are regularization parameter and the kernel parameter of SVM (SVM with RBF, q denotes the width of RBF function; SVM with Poly, q denotes the degree of polynomial). δ is the kernel parameter of KPCA and n is number of principle components.

5 Discussions and Conclusions

The results of this research suggest that the kernel-based models were able to map the underlying data structure relating to young and ageing populations. To date, the analysis of gait data is a challenging endeavour because of a handful of complicate features such as high-dimensionality, temporal dependence, high variability, and highly correlation and nonlinearity. To the machine classifier, these features could definitely deteriorate the generalization performance. So it is essential for good generalization performance while selecting a proper pre-processing of input variables.

Here, we adopted KPCA as a pre-processing tool for gait data, because KPCA not only condensed the representation of gait but also revealed clinically relevant information that would have been difficult to interpret from the original variables. In the terms of gait data, kernel PCA extracts a substantially larger number of nonlinear principal components, and therefore allows spreading the information regarding the data structure more widely giving a better opportunity to discard some of the eigendirections where the noisy part of data resides. In this study, we also noticed some application issues for KPCA. Generally, KPCA can ravel global structure in the data, preserving the entire sample space. It is useful for making general conclusions about the study population; however, we are often interested in localized structures, such as specific details of pathological gait deviations. This information resides in pockets of the samples space and can not be successfully uncovered with KPCA.

As for why we choose SVM, we mainly consider situation of statistical small size issue due to gait data collection is costly. The performance of SVM classifier is non-sensitive to the sample size and the dimensions. It eliminates many of the problems experienced with NN such as local minima and over fitting. In addition, its ability to produce stable and reproducible results makes it a good candidate for solving many classification problems.

In conclusion, the experimental results show that our method is able to extract the useful information of gait data effectively, enhancing the generalization ability of the classifier and improving the accuracy of gait pattern recognition. Our method has a great potential in the clinic applications of early diagnosis of gait diseases and evaluation of outcome of treatment and rehabilitation.

Acknowledgment

The work was supported by National Natural Science Foundation of China through grants No.60271025.

References

1. Begg, R.K., Sparrow, W.A.: Gait Characteristics of Young and Older Individuals Negotiating a Raised Surface: Implications for the Prevention of Falls. Journal of Gerontology: Medical Sciences 55A, (2000) 147–154
2. Nigg, B.M., Fisher, V., Ronsky, J.L.: Gait Characteristics as a Function of Age and Gender. Gait and Posture, Vol.2 (1994) 213–220
3. Winter, D.: The Biomechanics and Motor Control of Human Gait: Normal, Elderly and Pathological. University of Waterloo Press, Waterloo (1991)
4. Begg, R.K., Palaniswami, M., and Brendan: Support Vector Machines for Automated Gait Classification, IEEE Trans.Biomed. Eng., Vol. 52 (2005) 828–838
5. Barton, J.G., Lees, A.: An Application of Neural Networks for Distinguishing Gait Patterns on the Basis of Hip-knee Joint Angle Diagrams. Gait and Posture, Vol. 5 (1997) 28–33
6. Holzreiter, S.H., and Kohle, M.E.: Assessment of Gait Pattern Using Neural Networks. Biomech, J., Vol. 26 (1993) 645–651

7. Chau, T.: A Review of Analytical Techniques for Gait Data, Part1: Fuzzy, Statistical and Fractal Methods. Gait Posture, Vol. 13 (2001) 49–66
8. Scholkopf, B., Smola, A J., Muller, K-R: Nonlinear Component Analysis as a Kernel Eigenvalue Problem. Neural Computation, Vol. 10 (1998) 1299–1319
9. Vapnik, V. N.: Statistical Learning Theory. New York: John Wiley and Sons Inc (1998)
10. Gage, J.: Gait Analysis in Cerebral Palsy. Mac Keith Press, Londres (1991)
11. Giannini, S., Catani, F., Benedetti, M. G., et al: Gait Analysis Methodologies and Clinical Applications. Amsterdam: IOS Press (1994)
12. Ding, C.H.Q., Dubchak, I.: Multi-class Protein Fold Recognition Using Support Vector Machines and Neural Networks. Bioinformatics, Vol. 17 (2001) 349–358
13. Pang, C.C.C., Upton, A.R.M., Shine, G., et al, A comparison of Algorithms for Detection of Spikes in the Electroencephalogram. IEEE Transactions on Biomedical Engineering, Vol.50 (2003) 521–526

Interactive Color Planning System Based on MPEG-7 Visual Descriptors

Joonwhoan Lee[1], Eunjong Park[1], Sunghwan Kim[1], and Kyoungbae Eum[2]

[1] Division of Electronic and Information Eng., Chonbuk National University,
664-14 duckjin-dong Jeonju-city Chonbuk-do, 561-756 South Korea
{chlee, for0511, zkll5577}@chonbuk.ac.kr
[2] Dept. of Computer Science, Kunsan National University,
573-701 Miryong-dong Kunsan-city Chonbuk-do, 573-701 South Korea
{kbeum}@kunsan.ac.kr

Abstract. In this paper, an interactive color planning system is proposed. The core of the system is an emotion-based pattern retrieval that can be implemented with two data bases. In the system a user can issue an adjective word that represents the desired feeling to the system. Then the system retrieves the color combinations or representative color patterns from the knowledge base. Again when the user selects a color combination or representative color pattern, the system provides similar product patterns from the product DB according to MPEG-7 visual descriptors. Finally the user can choose the color patterns and apply them to the 3D VRML space. The decision support system can help users who do not have enough knowledge about color planning and can be used to promote e-business.

1 Introduction

The relation between color and human feeling is widely used in color planning including interior design, fashion design, commodity design, and urban design. Interactive color planning system assimilates human color designer's process and is a decision support system that helps users to make color design of a space.

In the color design, the designer can use the color image scales as published by Gobayashi or IRI Color Research Center, in which particular colors or color combinations are related and arranged with emotional adjectives.[1][2] Also, he (she) can rely on his (her) knowledge base to find a representative color and harmonious colors.

The decision support system for color design was proposed by Nakanishi, in which the principal color was selected by Gobayashi's color image scale and the harmonious colors were searched with the constraints according to Moon-Spencer's theory.[3][4] In the system, fuzzy set theory was widely used to represent the constraints to reduce the search space and process vague adjectives.

This paper proposes a new architecture to do the interactive color planning. In the proposed architecture, a knowledge base can be constructed according to the dictionary of color arrangements or representative color patterns. In the experiment, actually we use the dictionary from IRI color research center for the purpose.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 570 – 573, 2006.

Also, the architecture involves a data base of color patterns such as wall papers, carpets, furniture, and so on. Those color patterns are evaluated by similarity comparison with the color arrangements or representative patterns obtained from knowledge base and the patterns with higher scores are retrieved to the user. Eventually the user can choose a pattern among alternatives to apply it to objects in 3D space constructed with VRML. In the experiments, the system for interior design gives promising results.

2 Color Planning System

The proposed color planning system consists of 4 modules as shown in Fig. 1. The first one is retrieval module, which provides the corresponding product (color) patterns according to input adjective query such as “cool”, ”elegant”, “simple”, etc. The second one is 3-D navigation module with 3-D virtual objects, which makes user to apply the retrieved product (color) patterns to the virtual space. The third one is administer module that can be accessed in order to construct knowledge base and pattern DB. The last one is GUI subsystem to help user interaction with the planning system and to integrate the whole modules.

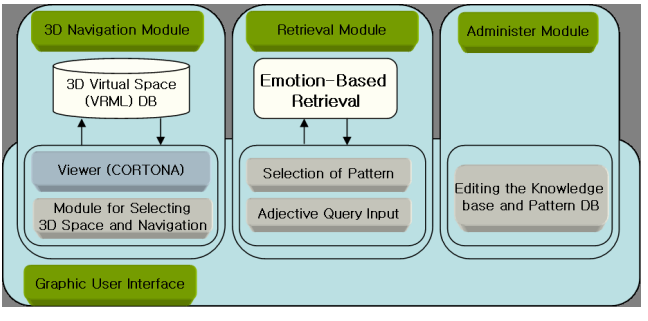


Fig. 1. Four subsystems of the color planning system

The core of the system is retrieval module. There are two data bases in it; the one is knowledge base and the other one is product (color) pattern DB. The formal stores the knowledge captured from experts of specific domain and the latter stores product patterns.

For an input adjective, the knowledge base recommends color arrangements to user through graphic interface. The knowledge base and the process to recommend color arrangements can be expressed as

$$KB\_C = F_c(adjective, color\_arrangements). \quad (1)$$

and

$$Recommended\ Color\_Arrangements = KB\_C(given\_adjective). \quad (2)$$

Also, the knowledge base can store representative product color patterns for emotion expressed by an adjective.

The other DB stores color products, color images or patterns. After the system recommends using the knowledge base, the user can interact with the system to select a color arrangement or a product pattern in order to give a key for product DB. The information retrieval in the product DB is similar to the search by example in conventional content-based image retrieval process. In other words, the similar product images or patterns to the selected key can be searched and retrieved in the process. In the product pattern retrievals, MPEG-7 visual descriptors and corresponding similarity measure can be widely used.

3 Implementation and Results of Experiment

In the system, the 3-D navigation module is constructed by Cortona viewer and 3-D objects are modeled VRML(Virtual Reality Markup Language) which makes user to navigate and apply the retrieved product (color) patterns. Administer module that can be accessed in order to construct knowledge base and pattern DB is implemented with a web server and browser. The GUI module to help user interaction with the planning system is also implemented as the administer module.

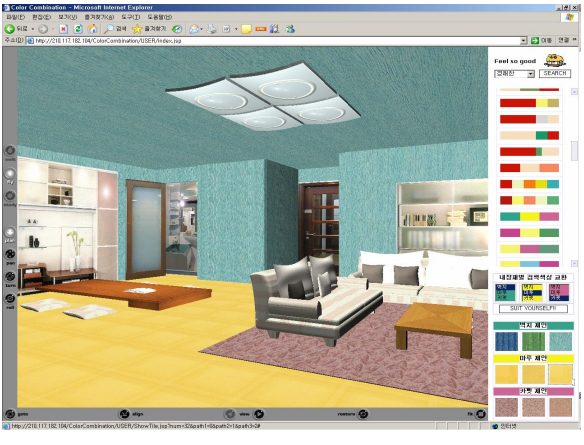


Fig. 2. Example of the system for “cheerful

After a user issue an adjective through the interface, the system recommends the color arrangements depending on the adjective. Once he (she) selects an arrangement, the system compares the arrangement with the colors of stored product patterns including wall paper, floor, and carpet. The comparison is done with the similarity measure of dominant color descriptor in MPEG-7. One can use various descriptors with similarity measures defined in MPEG-7 when the recommendation of the knowledge base is representative color patterns.

Some results of experiment are shown in Fig. 2. In the figures, the right part shows the input adjective and recommended color arrangements. The selected color arrangement

for interiors such as wall paper, floor and carpet are followed and finally the suggested product patterns for each interior component are displayed. The chosen products are applied to the 3D virtual space as shown in the left part of the figure. One can find the virtual space is pretty “cheerful” in the figure.

4 Conclusion and Further Study

In this paper, an interactive color planning system is proposed. The core of the system is an emotion-based pattern retrieval. The proposed architecture of emotion-based color pattern retrievals has an advantage that the conventional content-based image retrievals in MPEG-7 can be easily expandable by adding a proper knowledge base. One may construct emotion-based any information retrievals as same as the proposed architecture including music, sound, and video. Also, this type of architecture can be used to evaluate the emotion incurred from images and pictures.

There can be three variants of the emotion-based color pattern retrievals according to the application domain and added knowledge base. One of them was implemented for interior design with knowledge base of color arrangements. The experiment shows the promising results of decision support system that can help users who do not have enough knowledge about color planning and can be used to promote e-business.

There are some research problems to fully verify the proposed system with the architecture. The one is the performance evaluation of the results by human subjects. This requires a proper measure of evaluation with careful samples of subjects. Also, the other two variants and emotional evaluation of image and picture with knowledge base will be implemented sooner or later.

Acknowledgments. This research was supported by the Program for the Training of Graduate Students in Regional Innovation which was conducted by the Ministry of Commerce Industry and Energy of the Korean Government.

References

1. Gobayashi, *Color Image Scale*, Kohdansha Japan, 1990.
2. I.R.I Image Research Institute Inc, *Best Color Selection for Beautiful Design, Color Combination*, YoungJin.com, Seoul Korea, 2003.
3. Shohachiro Nakanishi, Toshiyuki Takagi and Satoshi Inamura, "Color Harmony by Fuzzy Set Theory,” *Proceedings of the Int. Conf. on Fuzzy Logic & Neural Networks*, Vol. 1, pp. 419-422, July, 1990.
4. Shohachiro NAKANISHI, Toshiyuki TAKAGI, and Taichi NISHIYAMA, "Color Planning by Fuzzy Set Theory,” *IEEE Int. Conf. on Fuzzy Systems*, Vol. 1, pp. 5-20, March. 1992.

Linear Program Algorithm for Estimating the Generalization Performance of SVM*

Dong Chun-xi, Rao Xian, Yang Shao-quan, and Wei Qing

School of Electronic Engineering, Xidian University, China
chxdong@mail.xidian.edu.cn

Abstract. A novel algorithm applied linear program to estimate the generalization performance of SVM is presented. When span is used to estimate the generalization performance of SVM, a series of quadratic programs needs to be solved, of which the object function defines an elliptic norm. Based on the theorem of convergence property of norm, the function can be approximated to an infinity norm, and then a linear program is achieved. The theoretic analysis and experiment results show that the method can estimate the generalization performance well and reduce the computation time greatly.

1 Support Vector Machines

SVM is a learning machine that classifies unknown sample into two classes. If samples are not linear separable, an inseparable SVM can be obtained by introducing a slack variable. It can be generalized to nonlinear case by introducing a kernel function that satisfies Mercer condition, which implements a nonlinear transformation from origin space (input space) to a high (maybe infinite) dimensional space (feature space), where the optimization classification hyper-plane is achieved. The procedure is easy to be implemented by replacing the inner produce with kernel function.

2 Using Span to Estimate the Generalization Performance of a SVM

Generalization performance is an important measurement of learning machines, it is also the basis of selecting parameters and features. In short, to estimate the generalization performance of a SVM is to estimate the potential error rate of a sample dataset with obtained SVM decision function. Loo method is a famous one. It is proved that the Loo error rate is an unbiased estimation of generalization performance.

Among these methods, Span method is a good method issued in recent years[3]. Besides the method described in this paper, most methods to estimate the generalization performance of SVM is an approximation to Loo method.

* Support by National Electronic Warfare Lab under grant no. 51435040105DZ0103.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 574 – 577, 2006.

3 Compute Span with the Linear Program

For convenience, the span can be rewritten with matrix form.

Let $\lambda = (\lambda_1, \cdots \lambda_{n^0})^T$, $\lambda_p = -1$, $H = (H_{ij}) = ((x_i, x_j))_{i,j=1}^n$, Span can be rewritten as:

$$S_p^2 = \min \lambda^T H \lambda, s.t. \lambda_p = -1, \sum_{i=1}^n \lambda_i = 0, 0 \le \alpha_i^0 + y_i y_p \alpha_p^0 \lambda_i \le C \quad (1)$$

An ellipse norm of λ is defined by (1)[4], its object function is convex. The QP defined by (8) is a convex quadratic program.

Lemma 2[5]**:** For a bounded convex program, Newton algorithm will stop after finite iteration, or it will produce an infinite point set $\{x^k\}$ with following characteristic:

(1). $\{f(x^k)\}$ is a strict monotonically decreasing sequence; (2). $\{x^k\}$ has a unique utmost point x^*, and $f(x^*)$ is the minimum of $f(x)$.

In fact, the characteristics are not limitation to Newton algorithm, all the convergence algorithms have the characteristics. If the procedure of searching the optimization point stops after finite iterations, an infinite point set can be construed with following manner: its first N points are same as that obtained by the procedure and the other points are same as the Nth point. The norm defined upon such set has following characteristic:

Lemma 3[4]**:** Necessary and sufficient condition that the sequence of vector $x^{(k)} = (\xi_1^{(k)}, \xi_2^{(k)}, \cdots, \xi_n^{(k)})$, $k = 1, 2, 3, \cdots$ in C^n converges to vector $x^{(k)} = (\xi_1, \xi_2, \cdots, \xi_n)$ is that the number sequence $\{\|x^{(k)} - x\|\}$ converges to zero with any norm $\|\cdot\|$.

According to lemma 1, 2 and 3, an issue can be achieved:

Issue 1: If the optimization procedure keeps unchanged, the optimization solution of (1) is same as that with ellipse norm.

The issue is easy to be proved with lemma 1, 2, and 3. However, it is inanition to search the optimization solution with ellipse norm and compute the value of object function with other norm. From another point of view, however, infinite norm is also a convex function, and if it is used to search the optimization solution, a monotonic convergent infinite sequence can be obtained. For the reason that a convex program has unique convergent point, if the procedure for searching optimization solution is long enough, the values of object function with different norm will be close enough. So the ellipse norm in (8) can be replaced with infinite norm and keep the error little enough, then the program (8) can be expressed as following:

$$S_p^2 = \min (\max |\lambda_i|), s.t. \lambda_p = -1, \sum_{i=1}^n \lambda_i = 0, 0 \le \alpha_i^0 + y_i y_p \alpha_p^0 \lambda_i \le C \quad (2)$$

Later experiment results will verify that the difference between the results of two object function defined by (1) and (2) is little enough in estimating the generalization performance.

In order to solve (2) with linear program, some further derivations are required. Firstly, λ_p corresponding to support vector x_p is excluded from solution procedure.

Secondly, after optimization, $\lambda_p = -1$ is compared with other λ_i to compute the span S_p^2. Finally, the $|\bullet|$ in object function of (2) will be removed. To do this, let $\lambda_i = R\gamma_i, R = \max|\lambda_i|$, then $-1 \le \gamma_i \le 1$, and (2) can be rewrite as:

$$\min R, s.t., \sum_{i=1}^{n} R\gamma_i = 1,\ 0 \le \alpha_i^0 + y_i y_p \alpha_p^0 R\gamma_i \le C; -1 \le \gamma_i \le 1; R > 0 \tag{3}$$

Finally, the program can be expressed as a standard optimization form:

$$\min\ -\sum_{j=1, j\ne p}^{n} \gamma_j, s.t. -\alpha_i \sum_{j=1, j\ne p}^{n} \gamma_j - y_i y_p \alpha_p \gamma_i \le 0, i \ne p, (\alpha_i - C) \sum_{j=1, j\ne p}^{n} \gamma_j + y_i y_p \alpha_p \gamma_i \le 0, -\sum_{j=1, j\ne p}^{n} \gamma_j < 0, -1 \le \gamma_i \le 1 \tag{4}$$

If $\gamma^0 = \{\gamma_i^0\}_{i=1, i\ne p}^{n}$ is denoted as the optimization solution of (11), $\lambda^0 = \{\lambda_i^0\}_{i=1}^{n}$ as that of (2), then: $\lambda_i^0 = R\gamma_i^0 = \frac{\gamma_i^0}{D} = \frac{\gamma_i^0}{\sum_{i=1, i\ne p}^{n} \gamma_i^0}$ and $S_p^2 = \max|\lambda_i^0| = \max\left(1, \sum_{i=1, i\ne p}^{n} \gamma_i^0\right)$.

4 Experiments

Experiment 1: The comparison of estimate results.

According to the features, the datasets selected from UCI repository [6] can be divided into some categories such as linearly separable or not, discrete or continuous, sparse or dense and so on. (1) Wisconsin Breast Cancer dataset (abbr. WBC). (2) IRIS dataset. The dataset is abbreviated as IRIS1vs23 and IRIS2vs3. (3) Tic Toc Toe dataset (abbr. TTT). (4) Prima Indian Diabetes dataset (abbr. PID).

The estimation results of generalization performance (error rate) are given in table 1; Loo rate (denoted as LooRate) and testing rate (denoted as ClassRate) is given in the table,LpRate is denoted as the result obtained with our method, and QpRate with origin one. The parameters of SVM are selected with the rule for minimizing the Loo rate. The last column is the result on USPS dataset, the detail result on USPS [7] dataset will be presented in the second experiment.

Table 1. Estimation results with linear program span, quadratic program span and Loo method

	IRIS2vs3	IRIS1vs23	WBC	PID	TTT	USPS
looRate	0.038462	0	0.04375	0.34839	0.015306	0.078864
ClassRate	0.081081	0	0.049808	0.3491	0.018373	0.1300
LpRate	0.0059172	0.033299	0.00050781	0.54543	0.0063515	0.09811
QpRate	0	0.032258	0	0.54194	0.005102	0.097792

From table 1, it can be seen that the results are very closed, it is to say our method can replace the origin one to estimate the generalization. At the same time, there is little error between the results obtained by span method and that by Loo method, the reason is that Loo method tests all the support vectors while span method tests

unbound support vectors only. Nevertheless it will not affect the validity of our method to replace the origin one.

Experiment 2: The comparison of estimation results and computation time.

In this experiment, the USPS handwrite optical character recognition dataset[7] is used. Table 2 is variation of computation time with parameters variations on different subsets.

Table 2. The comparison of computation time spent by two methods on different datasets

Gamma	0.001	0.002	0.004	0.0075	0.01	0.02	0.05
Qp on 3 subsets	8.954	50.953	380.42	3453.2	7334.2	42000	97955
Lp on 3 subsets	5.578	33.765	206.06	1633.6	4108	16368	26163
Qp on 4 subsets	14.515	98.625	1086.3	9378.9	22278	107870	356120
Lp on 4 subsets	9.235	67.532	446.39	6240.7	10933	24432	44889

From the results, it can be seen that either the estimation value or the varying trend is close. Further more, if more training samples are used, the difference will be more difficult to be distinguished; at the same time, computation time spent by our method is less than that by origin one, this advantage will become more clear when training sample number is large. The variation of computation time with the parameters can be explained as the change of the number of unbound support vectors.

5 Conclusion and Discussion

Span method is a good method issued in recent years, but span method needs to solve a series of quadratic program in solution procedure, its computation cost is huge. The method proposed uses linear program to compute the span, it simplifies the computation, shortens the computation time, and keeps the estimation accuracy unchanged almost. It is proved theoretically and experimentally that the method can take place of the origin one to compute the span and estimate the generalization performance.

References

1. V.N.Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 2000.
2. Christopher J.C.Burges, A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery, Vol. 2, No. 2, 1998, pp. 121-167.
3. Vladimir Vapnik, O. C., Bounds on Error Expectation for Support Vector Machines, Neural Computation, Vol. 12(a), 2000, pp. 2013-2036.
4. Cheng Yunpeng, Z. K. X. Z., Matrix theory (in Chinese), 2 ed., Publishing Company of Westnorth Technology University, Xi'an, 1999.
5. Chen Kai-zhou. Theory and method of optimization (in Chinese). Xidian University. 1998.
6. C.L.Blake and C.J.Merz, UCI Repository of machine learning databases. Irvine, CA: University of California, Department of Information and Computer Science [http://www.ics.uci.edu/~mlearn/MLRepository.html], 1998.
7. Available at http://www.kernel-machines.org/data/usps.mat.gz.

Solid Particle Measurement by Image Analysis

Weixing Wang[1] and BingCui[2]

[1] School of Electronic Engineering, University of Electronic Science and Technology of China, Post code: 610054 , Chengdu, China
wxwang@ee.uestc.edu.cn or znn525d@yahoo.com
[2] Department of computer science and technology, Chongqing University of Posts and telecommunications, Post code:400065, Chongqing, China
kxcbpro@163.com

Abstract. A good size measurement method should meet at least three criteria. These are rotational invariance, reproducibility and embody overall shape description (elongation / flakiness or angularity). According to these three criteria, this paper analyzes and evaluates several existing methods of solid particle measurement, such as Chord sizing, (multiple) Ferret diameter, equivalent circle, maximum diameter and equivalent ellipse etc. in image analysis. Based on the analyses and evaluations of the existing methods, we propose a new method - best-fit rectangle for size measurement that satisfactorily meets the criteria of rotational invariance, reproducibility and shape description.

1 Introduction

The methods of image analysis for measuring solid particles can be classified as one-dimensional (1D), two- or three- dimensional (2D or 3D). 1D is simple (i.e. chord sizing) and rough. 3D is complex and depends more on special equipments. 2D is widely used today, due to its accuracy (compared to 1D) and its industrial applicability (compared to 3D). 2D image analysis for measuring fragments has been applied in laboratories (where sieving analysis was and is used), analysis of rock muckpile, fragments on moving conveyor belts and on trucks etc. However, the common problem in the applications is how to set up a standard for obtaining fragment particle size distribution when using image analysis. There is no standard for image analysis yet, except that people try to make 2D size distribution from image analysis coincide with those obtained from sieving analysis. Therefore, a host of methods for converting 2D size distribution into sieving size distribution have been proposed and reported.

Currently, numerous fragmentation measuring systems have been developed and marketed, based on image-processing [1-20]. Generally, in 2D image analysis, a fragment particle's length, width, area and perimeter are measured. For evaluating the image analysis result, most researchers, directly or indirectly, convert 2D-size distribution into a size distribution weighted by volume because they believe the traditional method of sieving analysis, but there is little paper to discuss the reproducibility of the measured parameters.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 578 – 587, 2006.

To setup a standard of image analysis of rock fragmentation, the first important thing is to reasonably define measuring parameters and make measurement stable (reproducible), then to setup smart methods to convert visual parameters to engineering parameters. This paper will discuss the fragmentation measurement stability by using image analysis. In this study, several samples of railroad fragments have been used for measurement testing in a laboratory, and ten widely used image measurement parameters have been applied. In this paper, we mainly analyse the stability of image analyses for measuring fragment particles in the situations when (a) particles standing, (b) particles lying, (c) particles overlap and (d) are separated from each other. We also give some experimental results both from laboratory tests and field tests.

2 Widely Used Image Measurement Parameters

From a literature review, summarizing, the most widely used methods for solid particle size measurement appear: (1) Ferret diameter (length L, width W and maximum length $\sqrt{L^2+W^2}$); (2) Equivalent circle diameter (equivalent perimeter, radius and area of circle); (3) Equivalent ellipse (or rectangle) and (4) Maximum diameter (come from Multiple Ferret diameters or radiuses). Let us analyze and evaluate these widely used size definitions in image analysis.

The calliper diameter is the distance between two parallel tangents which are on opposite sides of the fragment. This method was proposed by L.R.Ferret 1931 [1], and the measurement is often referred to by his name. In systems employing boundary-coding techniques, a single pass around the boundary of an object noting maximum and minimum x and y co-ordinates will yield vertical and horizontal Ferret diameters (length L, width W, and maximum length $\sqrt{L^2+W^2}$).

The method of equivalent circle diameter has been widely used [3 - 11]. The term "Equivalent circle" has two different definitions. They are: (1) the equivalent circle of area and (2) the equivalent circle of perimeter (or mean radius, is a crude result of equivalent circle of perimeter). An equivalent circle is judged to have the same area or perimeter as the object to be measured, and the diameter of the circle taken as a measure of size. These definitions were proposed by Heywood (1964) [1].

The "mean radius" has been used by different image system to measure the size of fragment, the procedure is to count a number of radii, then average them. The general formula can be presented as

$$\bar{r} = \frac{1}{N}\sum_{i=1}^{N} r_i \tag{1}$$

where, $\bar{r}$ is mean radius, N is the number of sampled radii (r_i).

$$\text{when } N \to \infty,\ \bar{r} = \frac{1}{2\pi}\int_{-\pi}^{\pi} r(\theta)d\theta \tag{2}$$

From Eq.(2), it is clear that $\bar{r}$ is the radius of equivalent circle of perimeter.

Let us consider a solid particle's perimeter is P, area is A, the shape factor is defined as $Sh = \frac{P^2}{4\pi A}$ (≥ 1). The difference of radii (or diameters) between equivalent circle of perimeter r_p and equivalent circle of area r_a can be presented as

$$D_f = \frac{r_p - r_a}{r_a} = \frac{P}{2\sqrt{\pi A}} - 1 = \sqrt{Sh} - 1 \tag{3}$$

So D_f is the function of Sh, it is always greater or equal to zero. It is clear that r_p is always greater than r_a except for the solid particles with a shape of circle and non-roughness on solid particle's boundary, where r_p=r_a.

From the digitizing technique point view, the object area calculation is easier than object perimeter calculation; the later is more sensitive to the boundary roughness of solid particles. The method of the equivalent circle of area is easy to use. It is rotationally invariant, reproducible a roughness independent, but has no shape reflection. It is applicable for non-elongated objects.

For elongated objects, length and width are also measured by using Multiple Ferret diameter, within multiple Ferret rectangle boxes; the box with a maximum diameter is selected. This method is based on how many Ferret rectangle boxes are taken [12 - 18]. This kind of measurement includes some shape reflection comparing with the Chord sizing, Ferret diameter and Equivalent circle measurements. The number of the boxes will affect measurement result, which will cause the problems of rotational-invariance and reproducibility.

The multiple Ferret method, in the sense that $\max\{L_1, L_2, ..., L_D\}$ is chosen as the diameter, coincides with the dot product method, if new scanning directions are implemented using dot products with new coordinate axes. The only difference is that the directions chosen are multiples of 2, and multiple Ferrets are often not implemented as a one-pass algorithm.

For maximum radius measurement, after obtaining the gravity centre of a solid particle, the orientation is the axis through the centre and lies on the maximum length axis that also depends on how many radii are chosen. After obtaining the orientation of solid particles by utilizing the procedure, the lengths and widths of solid particles can be obtained by using radii in the decided orientations. In the method, only the length that is approximately close to maximum length can be obtained, the orientation depends on how many radii are used. The conditions of reproducibility, rotationally-invariant and shape reflection can not be fully met, because that the measured parameters of length, width and elongation are not stable.

The method assumes that every solid particle has a rectangular (or ellipse) shape with length L and width W, the L and W can be calculated based on equivalent perimeter and area [1]. The method is simple, but measuring result is more depending on solid particle's boundary roughness, can not gives out any shape information. It may be good for the solid particles with a similar shape and light roughness on the

boundaries. This method has the following procedure: a solid particle's area, gravity centre and least interior are obtained by using zeroth, first and second moments respectively, then a ellipse (or rectangle) that has the same zeroth, first and second moments as the solid particle has, is made, the length and width are used as the size parameters of the solid particle.

From the least moment concept, the equivalent ellipse and rectangle can be obtained as following description. Considering the length of major axis as 2α and length of minor axis as 2β, we obtain the minimum and maximum second moments of the ellipse about an axis through its centre: $\frac{\pi}{4}\alpha\beta^3$ and $\frac{\pi}{4}\beta\alpha^3$ respectively. For the equivalent rectangle, the minimum and maximum second moments are $\frac{4}{3}\alpha\beta^3$ and $\frac{4}{3}\beta\alpha^3$ respectively. This method can produce one group of parameters (length L, width W and ratio L/W) for every individual solid particle. It is rotationally-invariant because of least second moment theory applied, therefore, it also meet the condition of reproducibility and boundary roughness independence. For the shape reflection, from ratio L/W, the solid particle elongation can be known. The problem is that it can not distinguish angularity of solid particles.

3 Testing and Comparison Results

The material used in the study is railroad solid particles (crushed granite).Two solid particle samples (A and B) have been tested in a laboratory, the sample A consisted of 91 solid particles and the sample B consisted of 108 solid particles. The size range in sample A is with a sieve size between 31.5 and 50 mm, and the size range in the sample B is between 31.5-63 mm.

Images were taken of the solid particles in two different positions, lying and standing for the samples. For each of these positions 10 images were taken of the exact same solid particle sample. The solid particles were collected and replaced on the background material between each image.

The background material for the lying position was a luminous plastic sheet. The solid particles were placed randomly on the sheet in a stable position. No consideration was taken to which stable position it was placed; each solid particle may have several stable positions. Care was taken that the solid particles were not touching or overlapping in the field of view of the camera.

The background material for the standing position consisted of a layer of sand covered by a layer of luminous beads. The solid particles were placed; pushed down into the bed about half way, orient with their longest projected axis perpendicular to the background material. Care was taken that in the camera view the solid particles did not touch or overlap. A digital camera was mounted on a stand at 1.8 meters above the background material on which the sample was placed.

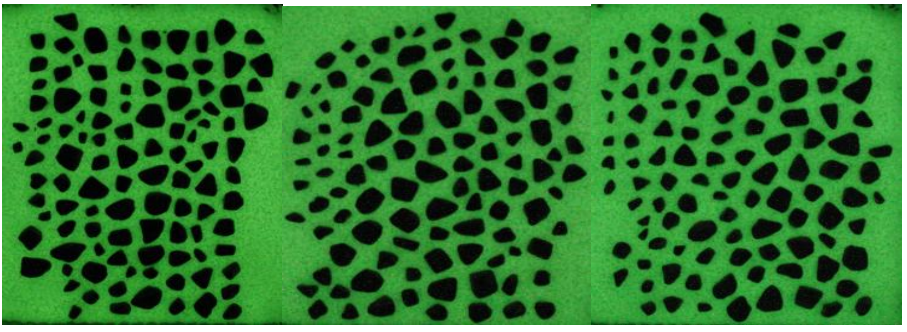


Fig. 1. Three different solid particle images where the same particles were in the position of standing

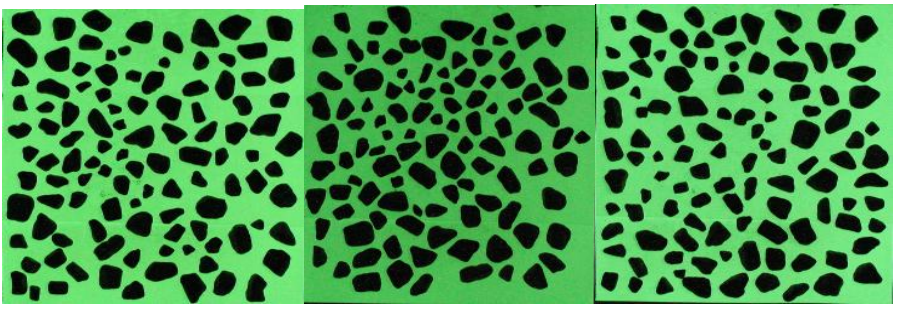


Fig. 2. Three different solid particle images where the same particles were in the position of lying

Image Analysis Method yields enormous amount of information. The object of this study has been to determine which of the parameter(s) yield the most reproducible results. The results show that in general there is good accuracy and reproducibility for most of the parameters; however, there are 4 parameters that yield substantial variation in the results.

There are some general trends for all the parameters: (1) for the parameters that show generally good accuracy and reproducible results, listed above, the mean standard deviation of the standard deviation as a percent of the mean is 2.1 %; (2) for the parameters that yield poor reproducible results the mean standard deviation with respect to the standard deviation in percent of the mean of the parameter is 43%; (3) there is a general trend for decrease of standard deviation with increase of size, this is not true for Perimeter and Roundness; (4) there are some erratic values, the smallest and the largest size layer in the sequence often shows extreme variation; (5) the solid particles in a lying position normally display lower standard deviation than do the solid particles in standing position; (6) the results for Ferret, Equivalent Ellipse, and Rectangle are nearly identical even though they are calculated in slightly different ways; (7) there is no uniform difference in standard deviation for the two resolutions, they often are nearly identical. There are however slight differences but there is no uniform trend than one of the resolutions yields less standard deviation than the other,

they switch relationships; (8) The variation in the standard deviation with respect to the mean of the 4 parameters with poor reproducibility clearly shows a strong relationships to size.

There are different reasons for this “diminishing tendency”. Overlapping might be only one of several main reasons. The factors which cause a diminishing tendency problem can be classified into: (1) camera system error; (2) the quality of aggregate materials; and (3) working ability of software (algorithm or program).

Camera system (image acquisition system) error includes mainly illumination conditions, camera location and quality. Poor illumination may cause the omission of parts of a particle. A good camera location of should make the image resolution constant over an image. Camera quality means the camera should not be affected by dust, rain and temperature etc.

The individual parameters (lengths and widths), all yield highly reproducible results. However the reproducibility decreases somewhat for the Elongation of these, ratio between the length and width. Noted also is that the results for the length between these three pair of parameters are nearly identical whereas they vary somewhat for the width. The Rectangle width is very slightly larger than Ferret and Ellipse widths. The definition of what is measured by these parameters is slightly different. One would expect that the Rectangle length might be slightly less than the Ellipse and Ferret length, this is not observed. As expect it is observed that the Rectangle width would be slightly greater than the Ellipse and Ferret width.

One would expect that the parameter Area would yield the least amount of variation since it is the area of the pixels represented in each solid particle and should not be rotational dependent. This is not the case. The lengths and widths of Ellipse, Rectangle and Ferret, and Equivalent Circle Diameter yield lower standard deviations than do area. Area has a mean standard deviation of the standard deviation is 4% where as the others all have means of less than 2%. It may be that the solid particles position in the field of view of the camera affects the results of Area more than it does the axial relationships of the geometric forms produced even if they are the equivalent ellipse and circle which by definition have the same Area as the solid particle.

Differences due to resolution are not observed for the parameters that yield good reproducibility of results. There is no difference in standard deviation between the two series of lying solid particles or the two series of standing solid particles; the sample A test with a resolution of 0.27 mm per pixel and after the sample B with a resolution of 0.63 mm per pixel. There is however a difference between the standard deviation for the two resolutions for the parameters of Minimum Radius, Radius Ratio, Perimeter and Roundness. In contrast to what might be expected the standard deviation increases for pictures taken with greater resolution.

There is a clear difference in standard deviation between the lying solid particles and the standing solid particles. In general the two series of lying solid particles have lower standard deviations than do the two series of standing solid particles. The variation is greater for the standing solid particles than it is for lying solid particles. This variation can be attributed to differences in the position of the solid particles. For both the lying and standing solid particles the position in the field of view of the camera should introduce some variation. This is assumed to be similar for both these positions. Thus the difference between these tow positions is how they are placed. For the lying solid particles variation is dependent upon which stable position they take. This

can change the viewed size substantially. For the standing solid particles the variation is due to how they are physically placed into the sand bed. The aim has been to orient the longest axis perpendicular to the background surface. This is done manually and human judgment can play an important role in the error introduced. Also the ability to actually orient the axis exactly perpendicular can be difficult to obtain. Size variation would be introduced if the axis were slightly dipping. Variation could also be due to which end of the solid particle is placed up or down. And furthermore how deep down into the sand bed the solid particles were pushed. This would especially affect long solid particles placed on the outer edge of the view of the camera.

Perimeter and Roundness are parameters that would have described changes in the surface texture. These however do not produce reproducible results and are not reliable. Area, Equivalent Circle, and Maximum Radius show that the solid particles become smaller but are not easily related to the nature of the size change. One can see that the solid particles in the lying position decreased much more in size than did those in the standing position.

Table 1. The order of image measurement parameter reproducibility

No	Parameters	Mean variation	Maximum variation
1	Maximum Radius	1.1-1.2	0.6-7.7
2	Rectangle Length	1.1-2.3	0.3-10.5
3	Ferret Length	1.1-2.3	0.3-10.5
4	Ellipse Length	1.1-2.3	0.3-10.5
5	Equivalent Circle Diameter	1.7-2.5(2.1)	
6	Ellipse Width	1.7-2.3	0.5-10
7	Ferret Width	1.7-2.3	0.5-10
8	Rectangle Width	1.7-2.3	0.5-10

The axial dimensions are represented by many parameters; the lengths and widths of Ferret, Equivalent Ellipse, and the Rectangle. The three dimensions of the solid particles, longest, intermediate and shortest axis can be observed. The longest axis is observed in the lying position by these parameter's lengths. The intermediate dimension is represented both by the width of lying solid particles and the length of standing solid particles. The shortest axis is the width of the standing solid particles.

Based on the statistics for the variability related to the mean of the different parameters (table 1) it is possible to rank them with respect to reproducibility of results. The following parameters yielded the most reproducible results, in order with the best at the top:

In order to analyze the difference and relationship between overlapping and non-overlapping of solid particles, we took one sample of 306 solid particle particles and a sieving size range from 11.2 mm to 20 mm. The particles were measured by image analysis.

In image analysis, particles were put on a plane with a black or white color contrasting with the color of particles, and separated or overlapped by hand to make sure that every particle was in a stable attitude. The camera was installed above the plane vertically. A good illumination was controlled by light sources. The equivalent circle diameter (Circle) was obtained for each of the solid particles.

Anyway, there is a certain difference: the range is up to 8%, as shown in Fig. 3.

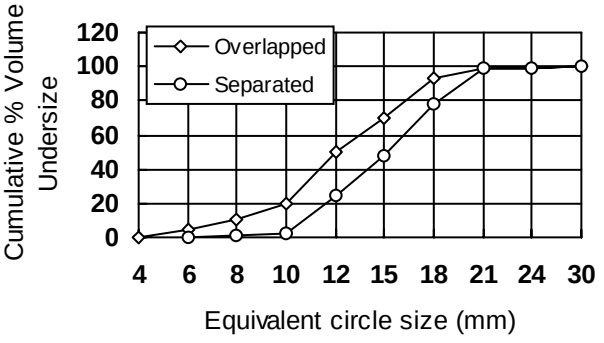


Fig. 3. Different sizes of a same material between overlapped and separated particles

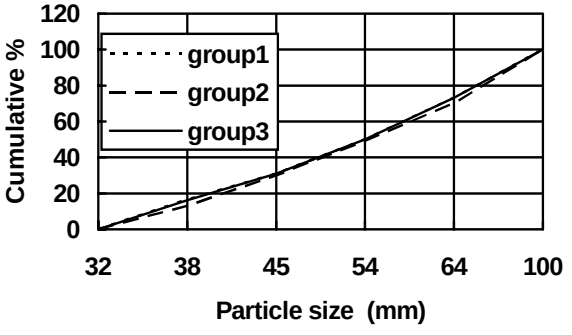


Fig. 4. Test in Jordbro, group 1,2,3 = image group 1,2,3.

The quality of solid particles mainly includes color variation on the surface of particles, degree of overlapping that relates to particle size distribution, and three dimensional geometry properties of particles. A large variation of color (grey value) can cause an over-segmentation problem. In size distribution, the narrower the sizes ranges, the higher the degree of overlapping. Complex three dimensional geometric properties and roughness on the surface of particles also create an over-segmentation problem.

The image analysis system can process 32-48 images as one processing group to provide a size distribution (consisting of 700 - 2000 particles) on average per three minutes. We kept the system working for almost 30 days. One hundred size distributions from image analysis have been picked out from our data base for analyses. Figure 4 shows one of examples of the size distribution on three consequently times. There is little difference between the three curves.

4 Conclusions

There are four parameters that yield extreme variation in results, Perimeter, Roundness, Minimum Radius and Radius Ratio. This should be clear before drawing conclusions

from these parameters. The reason for the high variation is not understood however they all seem to be related to size. It is more difficult for the program to determine the Minimum Diameter of small solid particles than it is for large solid particles. In contrast the standard deviations of Perimeter and Roundness increase with increased solid particle size. The larger the solid particles are the more variation there is in the length of the Perimeter and the degree of Roundness. The variations for these parameters is so great that it can not be explained by distortion in the view of the camera or change in the orientation of the solid particles in the 10 pictures; it must be due to how the program determines these parameters.

Acknowledgment

This contribution is part of the SCST (Science and Technology Bureau of Sichuan Province) project, Project No.05JY029-070.

References

1. Wang, W.X. and Fernlund, J., Shape Analysis of Aggregates. KTH-BALLAST Report no. 2, KTH, Stockholm, Sweden (1994).
2. Havermann, T. and Vogt, W., TUCIPS - A system for the estimation of solid particleation after production blasts, Measurement of Blast Solid particleation, Ed by J.A. Franklin and T. Katsabanis. Rotterdam: Balkema (1996), pp. 67-71.
3. Wang, Y., Zeng, S. and Wang, W.X., Study on the image analysis system of photographic film. Proc. for third period of Sino-Swedish joint research on science and technology of metallurgy, Publishing House of Metallurgical Industry, China (1992), pp. 116-126.
4. Montoro, J..J. and Gonzalez, E., New analytical techniques to evaluate solid particleation based on image analysis by computer methods. 4th Int. Symp. Rock Solid particleation by Blasting, Keystone, Vienna, Austria (1993), pp. 309 -316.
5. McDermott, C., Hunter, G.C. and Miles, N.J., The application of image analysis to the measurement of blast solid particleation. Symp. Surface Mining-Future Concepts, Nottingham University, Marylebone Press, Manchester (1989), pp. 103-108.
6. Ord, A., Real time image analysis of size and shape distributions of rock solid particles. Proc. Aust. Int. Min. Metall., 294, 1 (1989).
7. Lin, C.L and Miller, J.D., The Development of a PC Image-Based On-line Particle Size Analyzer. Minerals & Metallurgical Processing, No. 2 (1993) 29-35.
8. Von Hodenberg**,** M., Monitoring crusher gap adjustment using an automatic particle size analyzer, Mineral process, Vol. 37, (1996) 432-437 (Germany).
9. HAVER CPA, Computerized particle analyzer, User's Guide :CPA - 3, Printed by HAVER & BOECKER, Drahtweberei, Ennigerloher Straβe 64, D-59302 OELDE Westfalen, (1996).
10. Blot, G. and Nissoux, J.-L., Les nouveaux outils de controle de la granulométrie et de la forme (New tools for controlling size and shape of aggregates), *Mines et carrières, Les Techniques* V/94, SUPPL.-Déc., Vol. 96, (1994) (France).
11. J. Schleifer and B. Tessier, Solid particleation Assessment using the FragScan System: Quality of a Blast, Fragblast, Volume 6, Numbers 3-4 / December 2002, pp. 321 - 331, Publisher: Taylor & Francis.

12. Donald, C. and Kettunen, B.E., On-line size analysis for the measurement of blast solid particleation. Measurement of Blast Solid particleation, Ed by J.A. Franklin and T. Katsabanis. Rotterdam: Balkema, (1996) 175-177.
13. Norbert H. Maerz and Wei Zhou , Calibration of optical digital solid particleation measuring systems, Fragblast, Volume 4, Number 2 / June 2000, pp. 126 - 138, Publisher: Taylor & Francis.
14. Norbert H. Maerz [A1] and Tom W. Palangio [A1], Post-Muckpile, Pre-Primary Crusher, Automated Optical Blast Solid particleation Sizing, Fragblast,, Volume 8, Number 2 / June, 2004, pp. 119 – 136, Publisher: Taylor & Francis.
15. Paley, N., Lyman, G..J. and Kavetsky, A., Optical blast solid particleation assessment, Proc. 3rd Int. Symp. Rock Solid particleation by Blasting, Australian IMM, Brisbane, Australia (1990), pp. 291-301.
16. Wu, X., and Kemeny, J.M., A segmentation method for multiconnected particle delineation, Proc. of the IEEE Workshop on Applications of Computer vision, IEEE Computer Society Press, Los Alamitos, CA (1992) 240-247.
17. J. Kemeny, E. Mofya, R. Kaunda, P. Lever, Improvements in Blast Solid particleation Models Using Digital Image Processing, Fragblast, Volume 6, Numbers 3-4 / December 2002, pp. 311 - 320, Publisher: Taylor & Francis.
18. Kemeny, J., A practical technique for determining the size distribution of blasted benches, waste dumps, and heap-leach sites, Mining Engineering, Vol. 46, No. 11 (1994), 1281-1284.

Investigation on Reciprocating Engine Condition Classification by Using Wavelet Packet Hilbert Spectrum

Hongkun Li, Xiaojiang Ma, Hongying Hu, and Quanmin Ren

Key Laboratory for Precision and Non-traditional Machining Technology of Ministry of Education, Dalian University of Technology, P.R. China 116024
l.hongkun@163.com

Abstract. Nowadays, empirical mode decomposition (EMD) and Hilbert spectrum (HS) have been broadly investigated on non-stationary and nonlinear signal processing, especially on vibration signal analysis. But as diesel engine vibration signal wide frequency band, it leads to this method not decompose intrinsic mode function (IMF) successfully. Therefore, the obtained IMF is less meaning. For a better recognition of diesel condition because of its wider frequency band, this paper uses wavelet packet as preprocessing for HS analysis. It can effectively reduce wide frequency band and noise interference. Thus, the developed method is named as Wavelet Packet Hilbert Spectrum (WPHS). Experimental data of a DI135 diesel engine with different fuel supply advanced angle is used to evaluate effectiveness of the developed methodology on diesel pattern recognition. According to the recognition result, it can be concluded that this approach is very promising for reciprocating engine condition classification and preventative maintenance.

1 Introduction

Due to the complexity of mechanical systems and variation of working conditions, diesel engines often operate in an off design condition which could lead to poor performance, heavy wearing and even a break down of engines. Preventative maintenance based on an operation condition is a proven strategy that minimises maintenance and increases engine availability. It is normally convenient to monitor vibration signal of a diesel engine to evaluate the engine working condition and predict the occurrence of faults without interrupting its operation because its vibration signal contains most of information about diesel engine condition. Such an approach can provide preventive maintenance and replacement of failure components at an optimum period of engine operation.

But diesel engine vibration is a very complicated phenomenon caused by impulsive forces of moving components and external forces, such as cylinder combustion pressure, valve vibration, impulsive injection pressure, initial force, which occur simultaneously with external load disturbances. The vibration signals from cylinder head and cylinder block are typical non-stationary and nonlinear. At the same time, it also contains much noise [1]. All these have much effect on the signal analysis and pattern recognition. Thus, the traditional Fourier transform analysis is not effective for diesel engine signal analysis. Nowadays, advanced signal analysis technology which is

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 588 – 597, 2006.

called "time-frequency distribution analysis" has been used widely in nondestructive evaluation (NDE) applications. Time-frequency distribution (TFD) contains both time and frequency information and is suitable for non-stationary and nonlinear signal analysis. This method has been broadly investigated and applied on signal processing since it was introduced. As for now, there are many TFD methods for non-stationary signal analysis, such as short-time Fourier transform or spectrum, Wigner-Ville distribution and wavelet analysis. But all these methods are developed from traditional Fourier transform, not very suitable for diesel vibration signal analysis because of its typical non-stationary characteristics [2].

In this paper, the most recently developed signal analysis technique, i.e. Hilbert spectrum (HS) analysis [3], is applied to diesel engine pattern recognition. HS is developed according to instantaneous frequency, which offers narrow TFD and is an ideal method for non-stationary signal analysis. It has been broadly investigated and applied to signal processing. But it also has some limitations for real vibration signal analysis. The signal decomposition for HS analysis is a self-adaptive process, where frequency band width and noise have a strong influence on real signal. This has let to, in this study, the application of wavelet packet method to pre-treat signals and to obtain the required frequency band and improve the signal-to-noise ratio for analysis. The purpose of using wavelet packet is to improve the representation of TFD. After pre-processed by the wavelet packet, the HS is more accurate and convenient for real diesel vibration signal analysis. Thus, the developed method is named as wavelet packet Hilbert spectrum (WPHS). Experimental data of a DI135 diesel fuel injection system of different conditions is used to evaluate the improved methodology for system pattern recognition and fault diagnosis. It can be concluded that this method is very promising for diesel engine preventative maintenance.

2 EMD and Hilbert Spectrum

2.1 Signal Decomposition

To get a meaningful Hilbert spectrum, a new signal processing method was introduced by Huang named as empirical mode decomposition [3]. In 2000, Professor Ma put forward a new signal analysis theory, named as local wave method [2] which is the development of EMD and HS. For an arbitrary time series, $X(t)$, its Hilbert transform $H(t)$ can always be expressed as:

$$H(t) = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{X(\tau)}{t-\tau}d\tau \tag{1}$$

With this definition, $X(t)$ and $H(t)$ form a complex conjugate number. As a result, an analytic signal, $Z(t)$ is written as,

$$Z(t) = X(t) + iH(t) = a(t)e^{i\theta(t)} \tag{2}$$

$$a(t) = \sqrt{X(t)^2 + H(t)^2}, \theta(t) = \arctan \frac{H(t)}{X(t)} \tag{3}$$

$a(t)$ is the instantaneous amplitude of $X(t)$, and $\theta(t)$ is the instantaneous phase of $X(t)$. One important property of the Hilbert transform is that if the signal $X(t)$ is mono-component, then the time derivative of instantaneous phase $\theta(t)$ will physically represent instantaneous frequency $\omega(t)$ of signal $X(t)$, as shown Equation (4).

$$\omega(t) = \frac{d\theta(t)}{dt} \tag{4}$$

Unfortunately, almost all signals in practical applications hardly belong to the mono-component but multi-component ones. To make the instantaneous frequency applicable, Huang developed a new signal decomposition method named as empirical mode decomposition (EMD), which is able to decompose a signal into several individual and nearly mono-component signals [3]. The mono-components obtained by EMD are named as intrinsic mode functions (IMF) to which the instantaneous frequency defined by Equation (4) can be applied. An IMF is a function that satisfies with the following two conditions: (i) in the whole data set, the number of extrema and the number of zero crossing points must either equal or differ at most by one; (ii) at any point, the mean value of the envelope defined by the local maxima and envelope defined the local minima is zero. Data satisfied with IMF will have well-behaved Hilbert transform, from which a meaningful instantaneous frequency can be calculated. The decomposition process can be found in reference. According to EMD, original data is decomposed to n intrinsic mode components $C_i(t)$ and a residual component r_n as Equation (5).

$$X(t) = \sum_{i=1}^{n} C_i(t) + r_n(t) \tag{5}$$

2.2 Hilbert Spectrum

After finding each IMF component, Equation (6) can be obtained. It gives both the amplitude and the frequency of the real part (RP) of each component as a function of time.

$$X(t) = \mathrm{RP}\sum_{j=1}^{n} a_j(t) e^{i\theta_j(t)} = \mathrm{RP}\sum_{j=1}^{n} a_j(t) e^{i\int \omega_j(t)t} \tag{6}$$

Both the amplitude and the instantaneous frequency can be represented in a three-dimensional plot, in which the amplitude can be contoured on a time-frequency plane. The TFD of the amplitude is expressed by the Hilbert spectrum, $H(\omega, t)$, as shown in Equation (7).

$$H(\omega,t) = \mathrm{RP}\sum_{j=1}^{n} a_j(t)e^{i\int \omega_j(t)t} \tag{7}$$

3 Wavelet Packet Hilbert Spectrum

HS is developed according to instantaneous frequency. It offers narrow TFD and is an ideal method for non-stationary signal analysis. Thus, it has been broadly applied on signal analysis and pattern area [1], [2], [4]. Although HS are convenient for signal analysis, there are also some limitations for EMD process. One is from the frequency bandwidth which is also close related to gathering frequency. The other is from noise interference [1]. If the original signal can not be decomposed into several meaningful IMFs, its HS and HMS will not clearly reflects the characteristics of signal [4],[5]. The real monitored diesel engine vibration signal, its frequency band is very wide, from 0-10000Hz and even more. Therefore, EMD process must be affected, so does the accuracy of HS analysis. At the same time, noise also has much effect on HS analysis for the real monitoring signal [1].

Thus, it is better to shorten the frequency band width and remove noise interference before using HS for signal analysis. Among many methods, wavelet provides an effective way to separate different frequency band for analysis. It can decompose signal into several child frequency band. At the same time, it is also a powerful filter. Wavelet is often used to low frequency band analysis. As wavelet technology development, wavelet packet is more powerful and suitable for all frequency band analysis compared with wavelet [1]. It can decompose signal in the whole frequency band. At the same time, noise can also be filtered. It is more flexible for wider frequency band signal analysis compared with wavelet. The difference for wavelet and wavelet packet process is shown in Fig.1. Wavelet packet transform (WPT) can be used as a preprocessing for HS analysis, especially for diesel engine medium and high frequency band vibration signal analysis. Thus, the developed method is named wavelet packet Hilbert spectrum (WPHS). It will be more suitable for non-stationary characteristics signal analysis and put forward an effective method for reciprocating engine condition classification.

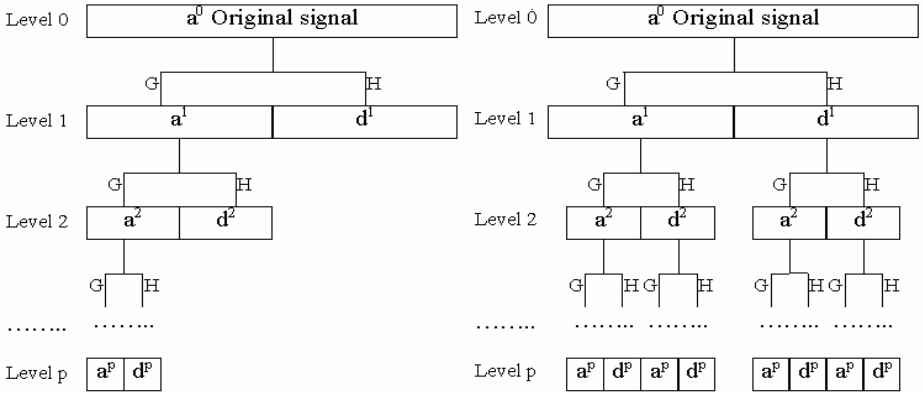


Fig. 1. Comparison of Wavelet and Wavelet packet decomposition process

4 Application of Wavelet Packet Hilbert Spectrum

4.1 Experiment

To test the effectiveness of this an approach on diesel engine pattern recognition, simulation experiments were carried out on a DI135 diesel engine, at Institute of Internal Combustion Engine, Dalian University of Technology. The main technical parameters of the experimental engine are shown in Table 1. The experiments were following three different diesel engine working conditions by changing its fuel supply advanced angle θ_{fs}. These are: reduced θ_{fs} (11deg), normal condition (15deg) and increased θ_{fs} (17deg). Other parameters of working condition are kept the same.

Table 1. Parameters of the DI135 diesel engine

Cylinder Diameter D (mm)	135
Piston stoke S(mm)	140
Piston swept volume Vs (L)	2.0
Compression ratioε	17.5
Declare power Pe (Kw)	14.7
Declare speed RPM(r/min)	1500
Injection nozzle	5×0.32×150
Injection starting pressure (MPa)	22

The acceleration transducer was located on the cylinder head, which is used to monitor the vibration signal. The top dead center (TDC) signal was also monitored together with cylinder head vibration. It is to determine the time information the vibration. It is very useful for feature extraction and fault diagnosis. The experimental condition is shown in Table 2.

Table 2. Experiment and sampling condition

Working speed RPM(r/min)	1500
Load	25%
Sampling Frequency (Hz)	25600
Analysis Frequency (Hz)	10000

4.2 Analysis Data Selection

Fig.2 shows the time domain vibration signal measured in one engine cycle under the increased θ_{fs} condition. For a single cylinder diesel engine working period, the vibration impulse is from four different sources: combustion process impulse response, exhaust valve close impulse response, throttle force impulse response and inlet valve close response. It is also obvious that diesel engine vibration signal is full of non-stationary characteristics. The combustion process impulse response is very different though the working period is successive shown in Fig.2 (b).

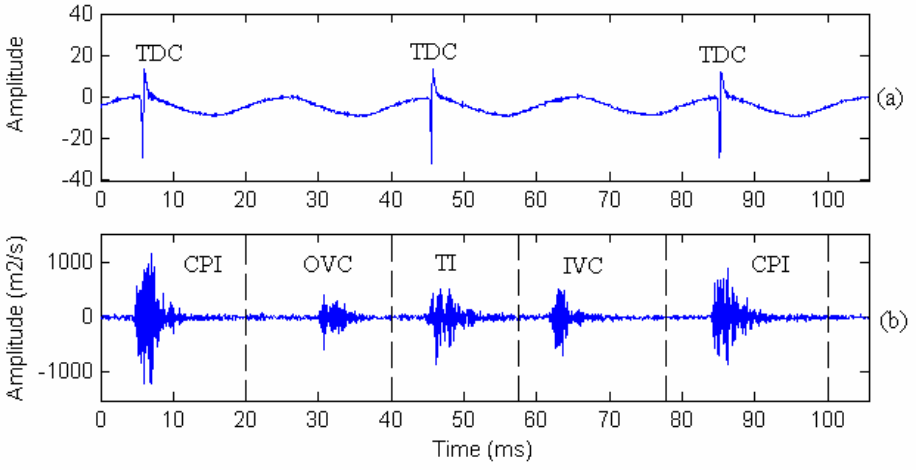


Fig. 2. Cylinder head vibration signal in a diesel engine working period. (a) TDC signal; (b) Vibration signal; CPI: combustion process impulse response; OVC: exhaust valve close impulse response; TI: throttle force impulse response; IVC: inlet valve close response.

The vibration of an engine cylinder head can be simplified as a linear system of multi-inputs and single output as shown. It contains the response of vibrations caused by combustion pressure impulsive response function, exhaust valve closing impulsive response function impact, throttle impulsive response function and inlet valve closing

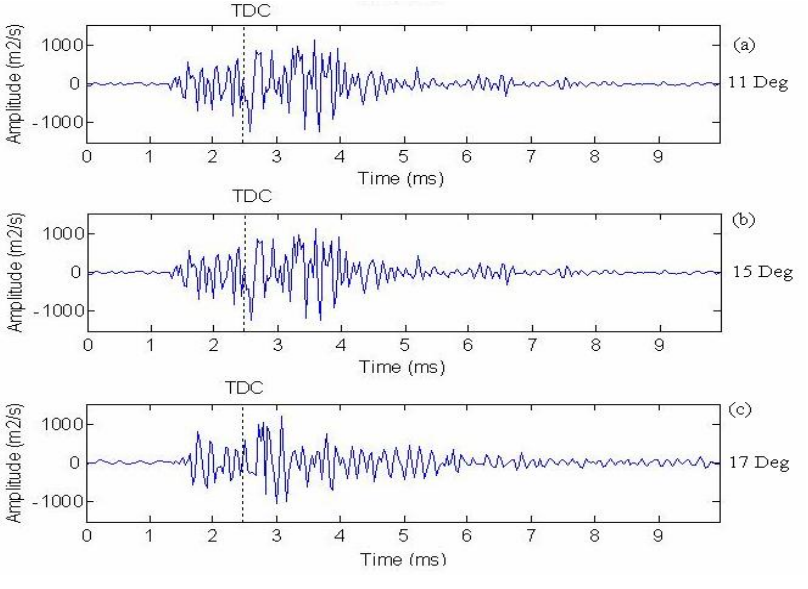


Fig. 3. Time domain signals of different working conditions

impulsive response function. At the same time, the vibration signals generated by impulse forces follow a fixed time regularity in every engine working cycle. Thus, the impulse response functions in the time domain can be separated and their individual characteristics studied [6]. In this investigation, the combustion impulse is used for pattern recognition, which can simplify the problem for analysis.

By using TDC signal for location determination, the data length of 256 point near the combustion process is taken from different condition for condition analysis shown in Fig.3. It can be found that max amplitude of the signal is basically the same. But the maximum corresponding time is different. Apart from the amplitude, it is difficult to distinguish the diesel engine condition because its vibration signal is full of non-stationary characteristics.

4.3 Hilbert Spectrum Analysis

According to the HS calculation, the HS of combustion process for different condition is obtain and shown in (i) of Fig.4, Fig.5 and Fig.6. The energy for reduced θ_{fs} (11deg) is mainly concentrated on high frequency band. It is about 8000-10000Hz. On the other hand, the energy for normal (15deg) is distributed on medium frequency band, which is about 4000-8000Hz. For the increased θ_{fs} (17deg), the frequency band is from 4000-10000Hz.

The vibration frequency band due to diesel engine combustion varies with the structure of the engine, particularly, the combustion chamber structure and the materials of the engine components which affect transmission speed of vibration. It changes with the diesel engine working condition. In normal condition, diesel engine vibration signal frequency is concentrated on medium frequency band. But diesel engine working condition can not be clearly demonstrated by the time and frequency information shown in HS of different conditions.

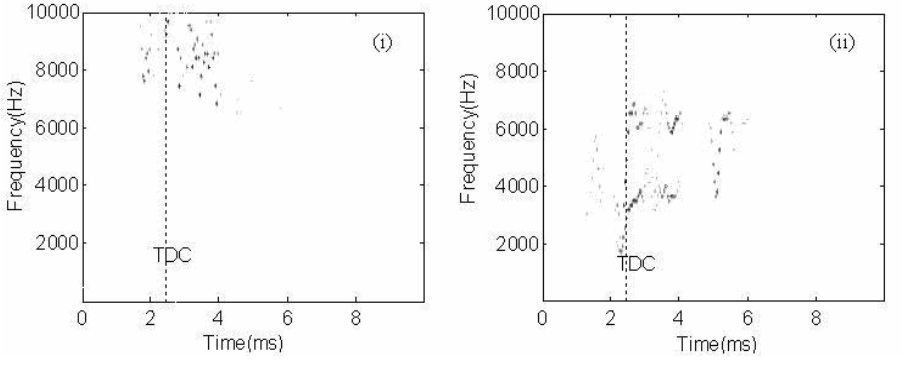


Fig. 4. HS and WPHS for reduced θ_{fs} 11Deg condition. (i) HS; (ii) WPHS

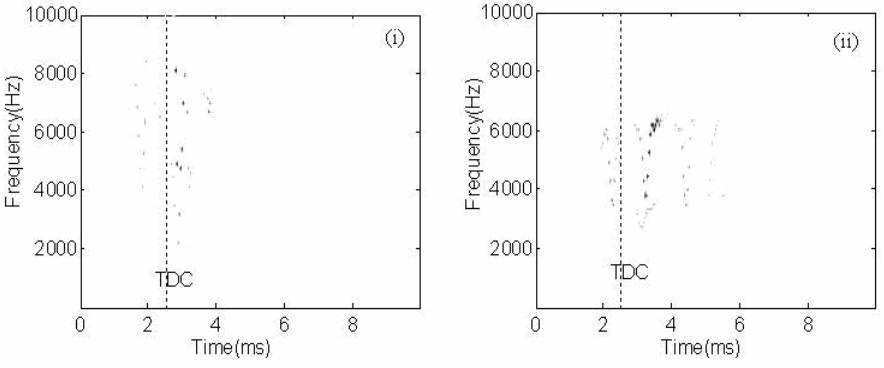


Fig. 5. HS and WPHS for normal θ_{fs} 15Deg condition. (i) HS; (ii)WPHS.

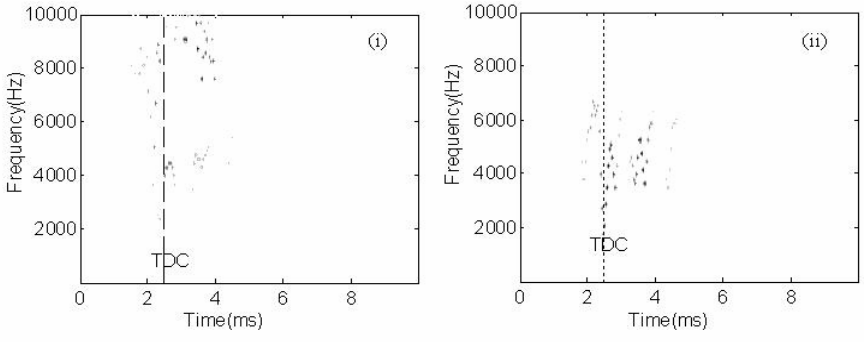


Fig. 6. HS and WPHS for increased θ_{fs} 17Deg condition. (i) HS; (ii) WPHS.

4.4 Wavelet Packet Hilbert Spectrum Analysis

As above analysis, it is better to analyse the medium frequency band feature for diesel condition analysis. Wavelet packet is used as the Hilbert spectrum analysis preprocessing. The 10^{th} order of orthogonal Daubechies wavelet basis is used for decomposition. In wavelet packet transform process, three levels decomposition have used. Then, the time domain signal of frequency band from 3750Hz to 6250Hz is reconstructed. WPHS for different conditions can be obtained according to the reconstructed time domain signal, shown in (ii) of Fig.4, Fig.5 and Fig.6 corresponding to different conditions.

Compared with different working condition by using WPHS, it is obvious that the durative time for combustion process reduces in the energy distribution plane with the fuel supply advanced angle increasing. It is unlike the HS, which is difficult to recognize. It can be concluded that the combustion time decrease with the increment of fuel supply angle. According to diesel engine working condition [7], the fuel supply angle is directly to the formation of mixture gas before ignition in the combustion chamber.

The bigger of the fuel supply advanced angle, the more mixture forms before ignition. After ignition, the combustion process will be very fast. The maximum pressure of combustion process will increase. Thus, it will lead to the improvement of vibration and reduction combustion time. At the same time, the exhaust air temperature also reduces since the reduction of combustion time and its energy distribution is more concentrated compared with small fuel supply advanced angle.

Diesel engine must work at the suitable θ_{fs}. Increased θ_{fs} will improve the cylinder pressure and the cylinder head vibration. On the contrary, reduced θ_{fs} will make the exhaust temperature and brake specific fuel consumption increase. Neither increased nor reduced θ_{fs} is good for diesel engine performance. It is a trade-off process. There is a best θ_{fs} for a determined diesel engine. It is more convenient to recognize diesel engine working condition compared with HS. Thus, it will put forward an effective tool for diesel engine pattern recognition and fault diagnosis.

5 Conclusions

In this paper, a developed approach named as wavelet packet Hilbert spectrum used on diesel engine condition monitoring and pattern recognition is demonstrated in detail. From the example of WPHS application, it can be concluded that this approach is very effective nondestructive evaluation for diesel engine pattern recognition and fault diagnosis. WPHS can effectively shorten the frequency and remove noise for signal analysis. It is more suitable for practical wide frequency band signal analysis. Thus, this promising method will contribute to the development of reciprocating engine condition monitoring and pattern recognition. Further research should also be carried out on its application.

Acknowledgement

The support from Chinese National Science Foundation (Grant No. 50475155) for this research is gratefully acknowledged.

References

1. Li H.K., Zhou P.L., Ma X.J..: Pattern Recognition on Diesel Engine Working Condition by Using a Novel Methodology - Hilbert Spectrum Entropy, Proceedings of the Institute of Marine Engineering, Science and Technology, (2005)43-48
2. Ma X. J., Yu B., Zhang Z.X.: A New Approach to Time-Frequency Analysis—Local Wave Method, Journal of Vibration Engineering, Vol.13(s), (2000)219-224
3. N.E. Huang, Z. Shen, S.R. Long, M.L.C. Wu, H.H. Shih, Q.N. Zheng, N.C. Yen, C.C. Tung, H.H. Liu.: The Empirical Mode Decomposition and The Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis, Proceedings Of The Royal Society Of London Series A- Mathematical Physical And Engineering Sciences Vol.454, (1998)903-995

4. Peng Z.K., Tse P.W., Chu F.L..: A Comparison Study of Improved Hilbert-Huang Transform and Wavelet Transform: Application to Fault Diagnosis for Rolling Bearing, Mechanical Systems and Signal Processing, Vol.19, (2005) 974-988.
5. Datig M., Schlurmann T..: Performance and Limitations of the Hilbert-Huang Transformation (HHT) with an Application to Irregular Water Waves, Ocean Engineering Vol. 31, (2004)1783-1834
6. Du H., and Shi X..: Reconstructing Cylinder Pressure from Vibration Signals Based on Radial Basis Function Networks. Proceedings of the Institution of Mechanical Engineers Part D- Journal of Automobile Engineering, Vol. 215,(2001)761-767
7. Zhou P.L., Li H.K., Xu F..: An Investigation into an Intensified Combustion Technology on DI Diesel Engine, 24th CIMAC, Tokyo(2004)

Research of a Novel Weak Speech Stream Detection Algorithm

Dong-hu Nie, Xue-yao Li, Ru-bo Zhang, and Dong Xu

College of Computer Science and Technology, Harbin Engineering University, Harbin City, Heilongjiang Province, China 150001
niedonghu@hrbeu.edu.cn

Abstract. Purpose of speech stream detection is to capture speech stream coming randomly in adverse acoustic environments. A novel robust method for speech stream detection is introduced based on both linear predict code all-pole model and lossless sound tube model to detect speech stream from inputs of wireless speech band communication. It makes use of autocorrelation distribution characteristics of variance sequence of linear predictive residual sequence to formulate two dimensions decision threshold vector. The decision threshold is adaptive to energy of background noise. It can make minimum decisions error. Plenty of signal stream data with various noises under various Signal-to-Noise Ratio and wireless speech band recordings on the spot were used to compare the proposed algorithm respectively with spectrum Entropy and short-time energy algorithm. The experiment results show that the new method for speech stream detection has good detection performance, and it performs well in adverse environments, and the speech stream detected sounds fluently.

1 Introduction

Speech Stream Detection(SSD) is a new research subject. Especially, weak speech stream detection under the background of complex and strong noise is a difficult problem. Its task is to capture speech presented random in endless acoustic stream. On the one hand, though Speech End Point Detection(SEPD) and Voice Active Detection (VAD) are also used for detecting speech, SSD's endless watching, unknown and protean acoustic environment and no cooperation with speaker etc. are traits differed from another two. On the other hand, it also stresses integrality of speech detected because information undetected can not be acquired again without cooperation with speaker. Book-phrase can not be loosed yet. That is not easy to be accomplished as sounds. For example, the pronunciation, “yes” or “no”, is no more than one second, but its meaning may be important. SEPD, VAD and SSD are all speech detection arithmetic. It is inevitable for them to deal with common problem including noise and real time etc.. SEPD is necessary part for improving the performance of speech recognition system. Since 1970s,many scholars have researched it and introduced their methods openly, such as Time-Frequency Power[1], Crosing-Zeros Rate[2], Entropy Function[3-5], Hidden Markov Model[6,7], Neural Network[8,9], Cepstrum[4,9], Short-Time Fractal[10], auto-correlation similarity distance[11], finite state machine [12] etc..

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 598 – 607, 2006.

Those methods can be classified into two kinds. One is to find robust new feature parameters for differing speech from noise. The other is to combine multi-features of speech. The two methods stress features of speech itself and can get good performance in SNR more than 5dB. But their limitations exist as follow: firstly, related information outside of speech itself can not be used, such as speech producing theory. Secondly, they do for specific language. Thirdly, examples for experiments are usually records under lab condition. Fourthly, it is only for specific background noise, so it is difficult for popularizing in practice. Fifthly, some of them need prior knowledge of speech and noise. VAD[13] is usually used in modern mobile communication and multimedia communication. It can detect speech in circuitry. It permits loss with context and can regain information with speaker's cooperation. Obviously, SE-PD and VAD can't satisfy with SSD's requirements. Currently, Shen Liran ect. makes use of method, such as High-Order Spectrum[14], Auditory Perception[15], Double Spectrum[16], Wavelet Transform[17] etc. to detect speech stream and get some achievement. But that is only just beginning of research work. Linear predictive coding (L-PC) technology is the kernal technology. This paper combins LPC with speech producing mechanism and applies them in speech stream detection. Its kernal method is to extract decision threshold by the autocorrelation coefficient distribution of Varivance of LPC Residual Sequence to detect speech in acoustic stream. In the process. It is key to make use of periodicity of LPC all-pole model's deconvolution. Experiment results show that the approach can detect weak speech efficiently under the background of strong noise. The ratio of detection can arrive more than 98% averagely.

2 Feature Extraction

2.1 Processing of LPC Residual

Speech Producing Model, shown in Fig.1, is based on both LPC all-pole model and lossless sound tube model. Its transfer function, H(z), that is

$$H(z)=\frac{S(z)}{U(z)}=\frac{G}{(1-\sum_{k=1}^{P}a_k z^{-k})} \tag{1}$$

where G is constant. $S(z)$ and $U(z)$ is respectively Z transform of input signal $s(n)$ and output signal $u(n)$. Relationship between $s(n)$ and $u(n)$ is given by the following expression.

$$s(n)=\sum_{k=1}^{P}a_k s(n-k)+Gu(n) \tag{2}$$

Let $\hat{s}(n)$ denote the predictive value at time n. It is can be approximated as a linear combination of the past P speech examples, such that

$$\hat{s}(n)=\sum_{k=1}^{P}a_k s(n-k) \tag{3}$$

If signal $s(n)$ can exactly satisfy with the model described by Eq.1, Eq.2 and Fig.1, we have the linear predictive residual sequence, $e(n)$, that is

$$e(n) = Gu(n) \tag{4}$$

Real signal isn't like that completely .But it can be estimated by Eq.5

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^{P} a_k S(n-k) \tag{5}$$

Where the sequence a_k, called as LPC coefficient, can be calculated by Eq.5 under the short-time average least-power-error criterion. Substitution of Eq.5 into Eq.4 yields $u(n)$'s expression, that is

$$u(n) = \frac{e(n)}{G} = \frac{s(n) - \sum_{k=1}^{P} a_k s(n-k)}{G} \tag{6}$$

In addition, if varied factors in process of producing speech can be considered adequately, LPC all-pole model can approach lossless sound tube model better. In this case, the characteristics of $s(n)$'s can be identified by analyzing the characteristics of $e(n)$'s. Based on speech knowledge that $e(n)$ is either quasiperiodic pulse train($s(n)$ is voiced sound) or a random white noise source($s(n)$ is unvoiced sound). Based on above basic idea, the speech stream detection method is proposed in this paper.

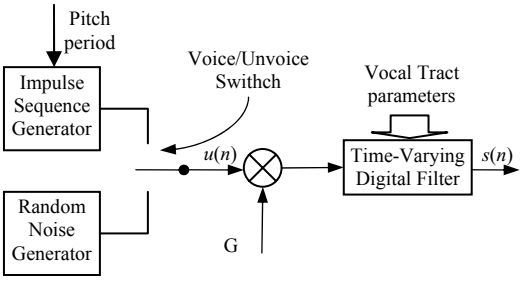


Fig. 1. Speech Producing model based on LPC model

To make clear the characteristics of $e(n)$'s, we calculate variance of $e(n)$, that is

$$D(l) = E\{[e_l(n) - E\{e_l(n)\}]^2\} \tag{7}$$

Where $D(l)$ is variance of the lth-frame residual sequence. Then, autocorrelating $D(l)$ yields auto-correlation coefficients of variance, $R(n)$,that is

$$R(n) = \sum_{l=0}^{M-n-1} D(l)D(n+l) \tag{8}$$

Where M denotes frame number.

2.2 Decision Threshold Feature Extraction Algorithm

Algorithm of threshold feature extraction is as follow, shown as Fig.2

- Signal Stream is weighted by Hamming window
- Band-pass filter consists of two low-pass filters
- High frequency pre-emphasis by using 1th-order FIR filter complements the effection due to glottal pulse and lip radiating.
- Calculating the variance of LPC residual sequence
- Calculating autocorrelation coefficients of variance sequence of LPC residual Sequence
- Making use of 1^{th} order auto-regress model to produce curve-fitting of autocorrelation coefficients
- Calculating slope of auto-regress curve and standard deviation of autocorrelation coefficients against auto-regress curve yields decision threshold for detection.

Operation of variance and auto-correlation reduces dimension, keeps short-time stationarity of speech and makes use of information of multi-frames. So it can improve system's distinguishability largely. Band filter gets rid of the industry frequency noise and high frequency components due to pulse noise.

3 Experiment Analysis and Result

3.1 Experiment Data

Experiment data is as follow:

- Pure speech data. 2002 Rich Transcription Broadcast News and conversational Telephone Speech, from Linguistic Data Consortium speech database(LDC), is about 30 minutes. Effective length of speech in it is about 15 minutes.
- Speech data with noise. Wireless speech band recordings from six frequency channels, including multi-language of Chinese, English, Russian, Korea, Japan etc., include complexed background noise. Signal-noise ratio is lower. Its length is about 34 hours and Effective length of speech is about 8.7 hours.
- Noise data. Noise data. From NOISEX-92 noise database, including speech babble noise, Volvo car interor noise, F16 cockpit noise, Tank noise, HF channel, factory floor noise, white noise.

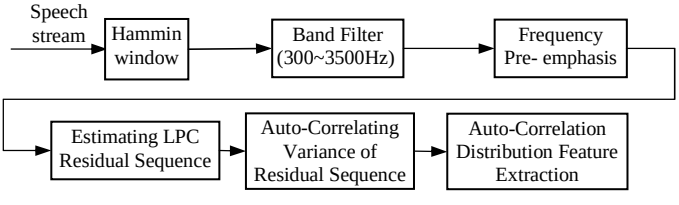


Fig. 2. Process of Decision threshold Feature Extraction

3.2 Experiment Analysis

Fig.3 shows the time domain waveform of speech and noise include pure speech from LDC, speech with noise from Short-wave field records, babble noise, Volvo car interor noise, F16 cockpit noise, tank noise, HF channel, factory floor noise and white noise from NOISEX-92. Fig. 4 shows autocorrelation coefficients of variance of LPC residual sequence for signals above. Observing Fig.4, we can get the following conclusions:

- The curves of various noise look relatively flat.
- The curves of pure speech and speech with noise change crookedly.
- Pure speech and speech with noise can be differentiated from others noise by a straight line.

Naturally, we will bethink of using auto-regress model to fit autocorrelation coefficients curve. That is

$$y(n) = a * x(n) + b \tag{9}$$

where $x(n)$ is x-coordinates of curve, $y(n)$ is y-coordinates of curve. At the rule of least squares error, we can solve the parameter a and b and yield 1^{th} order auto-regress equation, shown in Eq.9. Auto-regress curves of various signals above are shown in Fig.5. Seeing from the picture, we can know different slope among speech and noise. From Fig.4 and Fig.5, we can catch on different standard deviation of autocorrelation coefficients against curve-fitting.

In term of conclusions above, by calculating standard variance or slope, we can get appropriate two dimension threshold vector to distinguish speech from noise. That is, whether speech exists in signal stream or not, it can be judged by comparing it with the adaptive reference threshold vector. Let γ denote the reference threshold, we can calculate γ by Eq.10.

$$r = \begin{cases} \gamma_0 & E \le E_0 \\ \dfrac{\gamma_0 - \gamma_1}{E_0 - E_1} E + \gamma_0 - \dfrac{\gamma_0 - \gamma_1}{1 - E_1 / E_0} & E_0 < E < E_1 \\ \gamma_1 & E \ge E_1 \end{cases} \tag{10}$$

Where γ_1, γ_2 are the threshold of best SNR and worst SNR respectively. E_0, E_1 are background noise power respectively. Background noise power can be calculated by non-speech frame.

3.3 Detection Result for Various Environmental Conditions

To verify the effectiveness of the proposed algorithms, In the experiment, mixed signal with SNR from -5dB to 5 dB, by adding various noise from NOISEX-92 into pure speech from lDC, is used. Furthermore, we compared speech detection and denosing probability with spectrum entropy and short-time engery's method based on adaptive decision threshold.

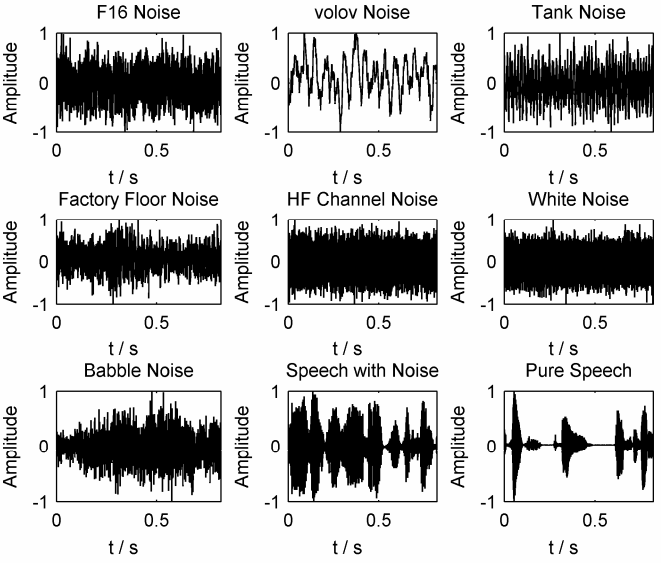


Fig. 3. Waveform of speech, speech with noise and several kinds of noise

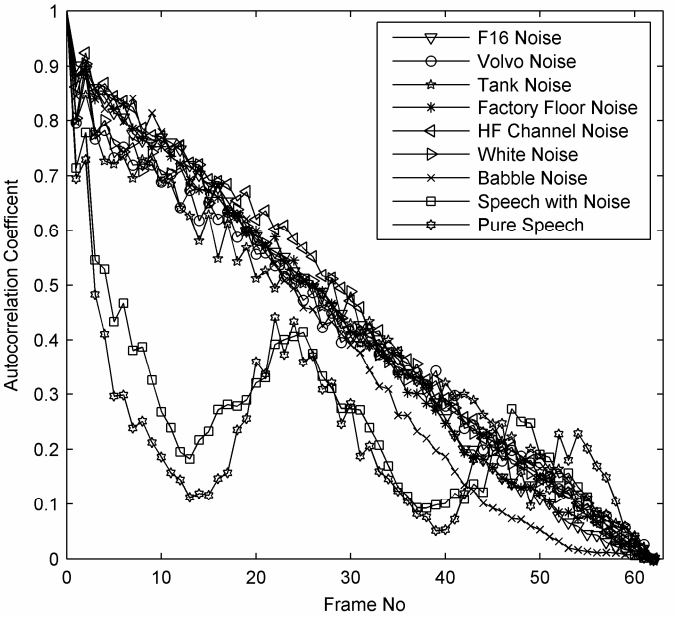


Fig. 4. Autocorrelation curve of varivance of LPC residual sequence for pure speech, speech with noise, and noise include speech babble noise, Volvo car interor noise, F16 cockpit noise, tank noise, HF channel, factory floor noise, White noise

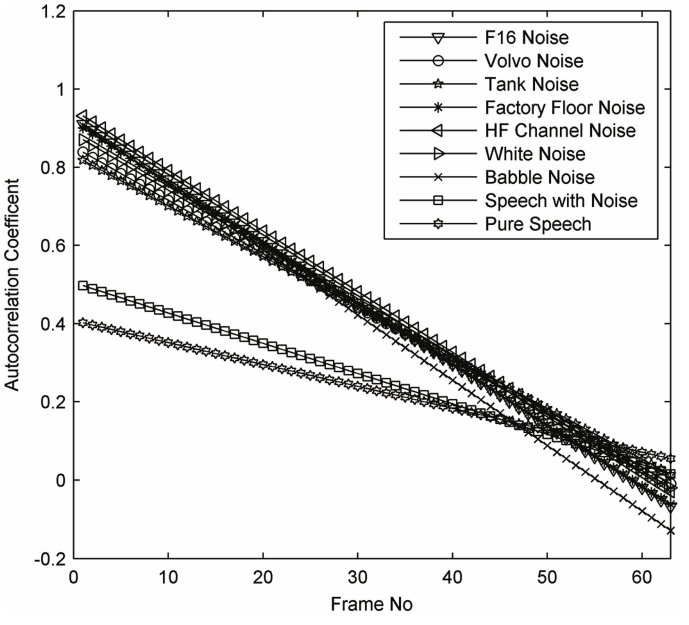


Fig. 5. 1th order auto-regress Curves of autocorrelation coefficients of LPC residual sequence of speech, speech with noise and several kinds of noise

Let P_s denote probability of speech detection, let P_n denote probability of denoising. That is

$$P_s=\text{SDL/ESL} \tag{11}$$

$$P_n=\text{DDL/(SSL-ESL)} \tag{12}$$

Where SDL is the length of speech detected, ESL is the length of effective speech in signal stream, DDL is the length of throwing off noise, SSL is the length of the whole signal stream. Experimental results are shown in Table.1.

Observing Table.1,the proposed arithmetic is obviously better than other two.In odB,detection performance is still satisfying.for babble noise, P_s and P_n are relativly lower,the reason is that babble noise listens like many people talking together in pub.

3.4 Detection Result of Six Channels Records

For furthermore evaluating detection performance,field records from six short-wave channels is used in the experiments.compared with spectrum entropy and short-time energy's algorithms,the results are given in table.2.

Observing Table 2, we can see the proposed algorithms' performance is also better than other two's under the condition of complicated background noise. The

probability of speech detection and probability of denosing are averagely more than 95%. Furthermore, in experiment, we find that speech detected by the proposed is smoother and more pellucid than the other two.

Table 1. Comparation of Detection Result for various Enviromental Conditions

Environments		Proposed SSD		Spectrum Entropy		Short-Time Energy	
Noise	SNR(dB)	P_s(%)	P_n(%)	P_s(%)	P_n(%)	P_s(%)	P_n(%)
Babble	5	98.32	94.69	92.07	62.61	88.89	98.16
	3	97.19	92.96	89.79	70.17	86.97	97.96
	0	94.55	86.83	87.30	76.96	87.78	97.35
	-3	92.43	85.34	87.73	78.78	91.83	79.04
	-5	91.08	82.78	85.31	73.91	87.30	88.43
Car Interior	5	99.69	97.84	95.30	99.98	95.08	98.86
	3	99.51	96.21	95.23	99.97	97.13	92.61
	0	99.67	95.98	89.64	99.94	94.32	73.91
	-3	99.68	94.70	85.46	99.27	93.18	73.04
	-5	98.70	93.13	78.91	99.71	87.35	71.18
F16-cockpit	5	99.51	98.17	89.12	64.34	86.27	98.31
	3	99.35	97.30	84.97	61.74	85.54	98.27
	0	99.19	96.47	84.11	77.71	81.47	88.10
	-3	96.79	94.26	81.83	60.87	61.74	88.78
	-5	94.86	92.96	82.65	42.61	50.22	90.81
White	5	99.75	99.23	96.78	89.12	92.17	95.23
	3	99.41	98.27	95.21	88.25	90.89	94.12
	0	99.24	98.01	92.24	84.07	85.34	87.13
	-3	97.31	96.32	83.12	76.45	79.23	85.34
	-5	96.24	93.12	80.19	72.17	77.61	80.17

Table 2. Detection Result of six Channels Speech Records

Data Records			Proposed SSD		Spectrum Entropy		Short-Time Energy	
Channel Number	Data Length(m)	Effective Speech(m)	P_s(%)	P_n(%)	P_s(%)	P_n(%)	P_s(%)	P_n(%)
1	210.05	91.73	98.17	98.35	80.22	78.86	84.07	86.87
2	458.64	21.31	95.76	98.26	71.32	73.65	67.53	63.77
3	496.25	203.78	98.83	99.98	89.31	81.72	88.74	88.46
4	175.70	68.42	94.63	95.34	81.67	78.91	78.07	78.55
5	559.90	98.10	96.76	97.64	86.31	84.92	80.01	83.65
6	146.63	41.22	98.97	99.91	90.45	92.42	88.37	85.77

4 Conclusion

Speech stream detection, is mainly used as intelligent speech detection in short-wave communication, is a meaning and challenging subject. Its difficulties are due to endless watching, unknown speaker, complicated acoustic environment and not to permit losing information. VAD and EPD arithmetics demanding SNR more than 5dB

determine that they are not directly used as speech stream detection for short-wave channel. In this study, basing on mechanism of speech producing model and LPC all-pole model, standard variance of autocorrelation coefficients of LPC predictive residual sequence is applied to differentiate speech from noise. Not only short-time stationarity can be guaranteed but also multi-frame information is made use of. In some range of field data, experiment gets good detection performance for SNR no more than -5dB.Of course, good system performance depends on exact adaptive threshold estimation which is difficult as well as noise estimation.

Acknowledgements

This study was supported by the National Nature Science Foundation of China (No.60475016).

References

1. Yie, T., Zuoying, W., Dajin L.: Robust word boundary detection through linear mapping of the sub-band energy in noisy environments. Journal of Tsinghua University (Science and Technology), Beijing, China 42 (2002) 953-956
2. Li Q., Zheng J., Zhou Q.: A robust, realtime endpoint detector with energy normalization for ASR in adverse environments. IEEE International Conference on Acoustics, Speech And Signal Processing (ICASSP 2001), Salt Lake City,USA1 (2001) 574–577.
3. Huang, L.S., Yang, C.H.: A novel approach to robust speech endpoint detection in car environments. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Istambul, Turkey (2003) 1751–1754
4. Jia, C., Xu, B.: An Improved Entropy–based Endpoint Detection Algorithm. International Conference on Spoken Language Processing (ICSLP 2002), Taipei (2002) 285–288
5. Wang, X., Qi, D., Bingxi,W.:A speech endpoint detector based on eigenspace-energy-entropy.Journal Of China Institute Of Communications 24 (2003) 125–132
6. Kosmides, E., Dermatas, E., Kokkinakis, G.: Stochastic endpoint detection in noisy speech. International Workshop on Speech and Computer(SPECOM97), ClujNapoca, Romania (1997) 109–114
7. Jie, Z., Xiaodong, W.: Speech Signal Endpoint Detection Method Based on HMM in Noise.Journal of Shanghai Jiaotong University,China 32 (1998) 14–16
8. Yuhong, L., Qiao, L., Qiang, R.: Speech signal endpoint detection and separation based on improved fuzzy ART. Systems Engineering and Electronic, China 26 (2004) 1151–1154
9. Guangrui, Hu., Xiaodong, Wei.: Endpoint Detection of Noisy Speech Based on cepstrum. Acta Electronica Sinica, China 28 (2000) 95–97
10. Yaqiang, S., Genliang, F.: Two End Points Detecting and Filtering on Low SNR Speech Signals Based on Short–time Fractal Dimension. Journal of Zhejiang Normal University (Natoral Sciences),China 22 (1999) 16–21
11. Feili, C., Jie, Z.: A New Method of Endpoint Detection Based on Distance of Auto-Correlated Similarity. Journal of Shanghai Jiaotong University, China 33 (1999) 1097–1099
12. Qiuan, H., BO, J., Bingwen, W.: Endpoint detection of Chinese digital speech based on finite state machine. Journal of Hubei University (Natural Science Edition), China 26 (2004) 35–38

13. Javier, R., Jose, C.: Efficient voice activity detection algorithms using long-term speech information [J]. Speech Communication, 42 (2004) 271–287
14. Liran, S., Xueyao,.: Speech stream detection based on higher-order statistics. International Conference on Machine Learning and Cybernetics, Xi'an, China 5 (2003) 3086–3089
15. Xueyao, L., Liran, S. A new method of robust detection for speech stream. Proceedings of 2002 International Conference on Machine Learning and Cybernetics, Beijing, China 5 (2002) 1066–1069
16. Liran, S., Xueyao, l. Speech stream detection based on one and half spectrum. Proceedings of the 5th World Congress on Intelligent Control and Automation, Hangzhou, China 5 (2004) 4223–4226
17. Rubo, Z., Jiashi, L., Xueyao, L., Liran, S.: Speech stream detection in non-Gaussian background noise based on statistic characteristics of wavelet coefficient. Journal of Harbin Engineering University, Harbin, China 25 (2004) 487–490

Large Diamond and Small Pentagon Search Patterns for Fast Motion Estimation*

Jianbin Song, Bo Li, Dong Jiang, and Caixia Wang

Digital Media Laboratory, School of Computer Science and Engineering,
Beihang University, Beijing, 100083, China
buaasjb@yahoo.com.cn, boli@buaa.edu.cn,
duff99@163.com, yueliangbuku@163.com

Abstract. In fast motion estimation, a search pattern with different shape or size has a very important impact on search speed and distortion performance. A motion estimation algorithm based on the novel large diamond and small pentagon search patterns is proposed in this paper. The stride of the proposed large diamond pattern is 3 pixels and it just need 2 or 3 search points for every new search step. So, the large diamond pattern can find the lager-motion vector quickly compare with the 2-pixel-stride hexagon pattern, and, it is does not easy to lose correct search path and fall into locally optimum point compare with the three-step search. The proposed small pentagon pattern can do more refined search than the small diamond pattern and small hexagon pattern. The proposed algorithm may find any motion vector regardless of no-, small-, medium- ,or large-motion with fewer search points than the diamond search algorithm and the hexagon-based algorithm while maintaining similar distortion performance. Experimental results substantially justify the further improvement achieved of the LDSPS algorithm compared with several other popular fast algorithms.

1 Introduction

Block-based motion estimation is an essential technique of most international coding standards such as MPEG–4, H.264 and so on. However, the computation of motion estimation is extraordinarily complex. If use the full search (FS) motion estimation, it approximately consumes 60% or higher of the encoding time. Fast motion estimation algorithm is always the hot research in the video compression field. The search pattern and search strategy lead a vital role in speedup of a fast algorithm. The three step search (TSS) [1] is extremely easy to miss the correct search path and not unable to find the best matching block. Comparing with TSS, the performance of new three step search (NTSS) [2] and four step search (FSS) [3]is improved greatly, but the zero-vector coding block also needs 17 search points. The block based gradient descent search (BBGDS) [4]carries on the search by the length of 3×3 stride and it is easy to fall into locally optimum point The diamond search (DS) [5,6] use diamond pattern and the recently proposed

* Supported by the NSFC(60505007), the Program for New Century Excellent Talents in University, the Doctoral Education Foundation of MOE and the ERIPKU. The research is made in the State Key Lab of Software Development Environment.

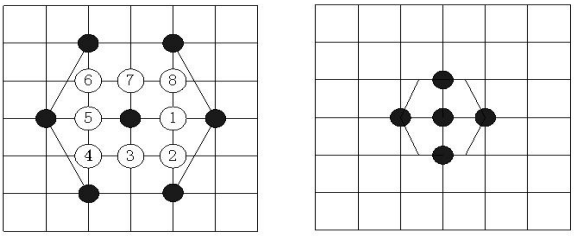
L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 608 – 616, 2006.

hexagon-based search algorithm(HEXBS) [7] use hexagon pattern respectively to remarkably decrease computational complexity of motion estimation, because of the short search stride and the more search points for every step, there exists redundancy of search and the performance is still waited for further enhancement.

Many efficient motion estimation motheds such as literature[8~9] based on good search pattern is proposed. So the search patterns have very important impact in fast motion estimation. This paper aims to study the search pattern and search strategy, doesn't study the prediction of search beginning point and the search termination criterion. A fast motion estimation algorithm based on novel large diamond and small pentagon search patterns are proposed, which further speed up of the motion estimation algorithm. Section 2 reviews and investigates HEXBS algorithm. Section 3 firstly explains the large diamond and small pentagon patterns, and then describes the steps of the new LDSPS algorithm. Section 4 compares the number of search points of HEXBS with LDSPS. Compare with FS, NTSS, FSS, DS and HEXBS, section 5 presents and analyzes the experiment results. Section 5 concludes this paper.

2 The HEXBS Algorithm

Compared with the DS algorithm, the computation speed of the HEXBS increases by 21.4% ~ 36.7%[7], while image quality nearly does not decrease. However, in Fig.1(a), we observe the pixels in the big hexagon, if the pixel 3 or 7 is the best matching position, it needs 11 search points; if the pixel 1 or 5 is the best, it needs two steps of big hexagon search, 14 search points possibly; if the pixel 2/4/6/8 is the best, it needs two steps of big hexagon search and altogether 14 search points. In addition, the search stride of the large hexagon is 2, so the HEXBS algorithm needs many steps to find the best position for the coding block which has large motion, search efficiency is still insufficient and needs to improve. Statistical results suggest that 40%~70%[2][5] best optimum point is around the center point, it has the center-biased characteristic. Therefore, the HEXBS algorithm still has search redundancy. Designing more reasonable search pattern and search strategy can evidently speed up the motion estimation.



(a) **(b)**

Fig. 1. Seach pattern of HEXBS

3 The LDSPS Algorithm

3.1 Search Pattern

Based on the proposed large diamond search pattern(LDS) and the small pentagon search pattern(SPS), this paper propose a fast motion estimation algorithm(LDSPS). may find any motion vector regardless of no-, small-, medium- ,or large-motion with fewer search points than the diamond search algorithm and the hexagon-based algorithm while maintaining similar distortion performance.

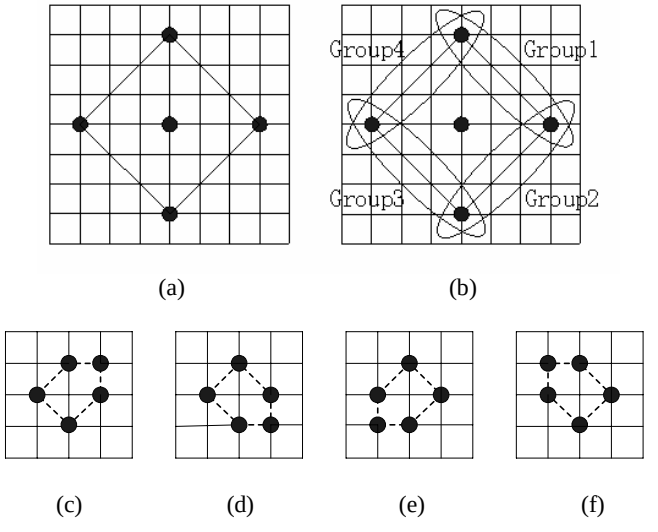


Fig. 2. Search pattern of LDSPS. (a) large diamond pattern; (b)group of horizontal and vertical MBD points; (c)~(f), small pentagon patterns.

1) LDS: LDS' search stride is 3 pixels and search 1 center point and 4 corner points at the first step and needs 2 or 3 additional search points for every new LDS step, whereas the stride of the DS is 2 pixels, at the first step it must search a center point, 4 corner points (vertex) and 4 edge points (face).According to different search directions, it needs 3 or 5 additional search points for every later step. In the HEXBS, the search stride of large hexagon is 2 pixels, the first step needs to search 1 central point and 6 vertexes, every later step needs 3 search points. If the best point is the adjacent pixel around the search center, LDSPS algorithm needs 1 step LDS, 1 step SPS and only searches 10 points. However, the famous HEXBS algorithm needs 11 or 14 points. The motion vector has the center-biased characteristic, therefore, for most motion block, the LDSPS algorithm can search the best point very quickly. For the movement who has large motion vector, LDSPS algorithm can find the best matching block in the reference frame quickly by using LDS because of its larger search stride. The larger the motion vector is, the more obvious the advantage of the LDS is.

2) SPS: SPS is appropriate for refined search. In this paper, minimum block distortion (MBD) is used for selection of the best position. SPS finds 4 horizontal or vertical points around the point which is selected as the best one in the last LDS. In order to improve search efficiency, a supplement search point is added. Here, we assume reasonably that the global minimum has a monotonic distortion and the nearer to the global minimum the smaller the distortion in all directions (horizontal, vertical or diagonal) within a small neighborhood around the global minimum. So, the selection of the supplement point is mainly based on the horizontal MBD point and the vertical MBD point in the previous LDS. In Fig.2(b), if the horizontal MBD point and the vertical MBD point compose Group1, the algorithm uses the small pentagon like Fig.2 (c), if Group2, uses Fig.2 (d), if Group3, uses Fig.2 (e), if Group4, uses Fig.2 (f). Comparing with the small diamond in DS and small hexagon in HEXBS, SPS finds five points and can achieve more refined result.

The search range of TSS, NTSS and FSS is restricted by the method itself, while the LDSPS algorithm proposed in this paper doesn't have the restriction, so it can find the best matching block in arbitrary search scope. The video encoder always confines the search scope considering the balance of speed and efficiency, such as ± 7, ± 15, ± 31 and so on. In this kind of situation, it just needs to add boundary judgment in the LDSPS algorithm. In section 5, the search window size of the experiment is ± 15.

3.2 Algorithm Process

The LDSPS algorithm can be divided into three steps as follows:

Step 1: Taking the search beginning point as the diamond center and starting first LDS. If the best position is the center point, then go to step 3, otherwise go to step 2;
Step 2: Taking the MBD point of the previous step as the center point, start a new LDS, if new MBD position of current step is the diamond's center point, then go to step 3, otherwise repeat this step;
Step 3: Judging the horizontal and vertical MBD point in previous LDS, if the horizontal MBD point and the vertical MBD point compose Group1, the algorithm uses the small pentagons like Fig.2 (c), if Group2, uses Fig.2 (d), if Group3, uses Fig.2 (e), if Group4, uses Fig.2 (f).After this step, the best position of motion estimation is obtained. After selecting the small pentagon, do the SPS, then the search process is stop.

A search path example is showed in Fig.3. There are 6 steps of LDS. The first step searches 5 points. The second, third and fourth step searches along the horizontal direction and each step needs 3 points. The fifth step searches along the vertical direction, adds 2 search points. The sixth step changes to the horizontal direction, needs 2 search points. The MBD point of this step is the center point of the diamond, so the LDS is stopped and the SPS is started. The horizontal and vertical MBD points compose Group4, therefore, pentagon like Fig.2 (f)is selected as final search pattern. This search example altogether needs 23 search points.

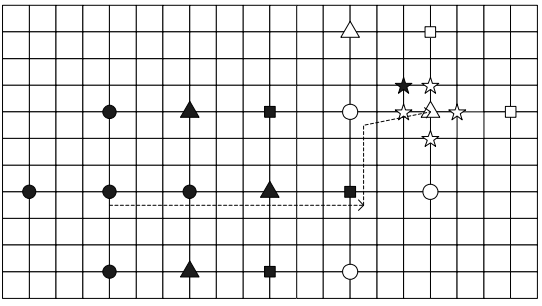


Fig. 3. Search path example

4 Search Points Contrast

The hexagon'center point and the large diamond'center point are the search beginning points. Supposing that the point around of the center within 4×4 search window is the best position respectively, the HEXBS and the LDSPS respectively calculate the fewest search points number and the total number is recorded in the Fig.4(a) and Fig.4(b). Fig.4 (a) shows the search number of different points as the best position using the HEXBS algorithm, Fig.4 (b) records the search point number of the LDSPS algorithm. Fig.4(c) shows the number of search points saved by LDSPS compared with HEXBS for each motion vector location. If the point dyed by deep color is the best position, it needs 14 search points in the Fig.4 (a) while the same position in Fig.4 (b) just needs 10 points, 4 search points are saved by using the LDSPS. As for the other points in the Fig.4, the LDSPS algorithm can saves 1 search point at least and 5 at most. Not only the large motion vector but also the small motion vector, LDSPS can improve efficiency of motion estimation all the time.

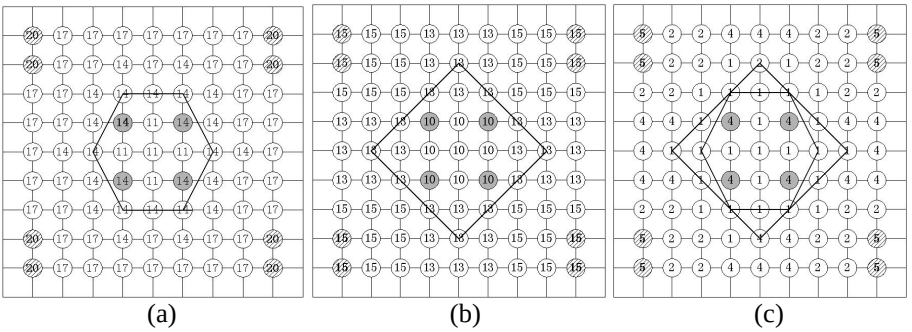


Fig. 4. (a) Minimum possible number of search points for each motion vector location by HEXBS algorithm. (b) Minimum possible number of search points for each motion vector location by proposed LDSPS algorithm. (c) Number of search points saved by LDSPS compared with HEXBS for each motion vector location.

In the video, the majority movement is along the horizontal direction, so this paper discusses the search points of the HEXBS algorithm and the LDSPS algorithm in the horizontal direction just as an example. Showed in Fig.5, the point (0,6k) is selected as the comparison point, H_k represents the number of search points when using the HEXBS, while N_k represents the number of search points when using the LDSPS. According to HEXBS and LDSPS:

H_0 = 7 + 4, altogether needs 2 steps, 11 search points;

N_0 = 5 + 5, altogether needs 2 steps, 10 search points;

H_1 = 7 + 3 + 3 + 3 + 4, altogether needs 5 steps, 20 search points;

N_1 = 5 + 3+3+5, altogether needs 4 steps, 16 search points;

H_2 = 7 + 3 + 3 + 3 + 3 + 3 + 3 + 4, altogether needs 8 steps, 29 search points;

N_2 = 5 + 3+3+3+3 + 5, altogether needs 6 steps, 22 search points;

H_k = 7 + 9k + 4, altogether needs 3k + 2 steps, 9k + 11 search points;

N_k = 5 + 6k + 5, altogether needs 2k + 2 steps, 6k + 10 search points;

The speed improvement ratio(SIR) is defined as follow:

$$\mathrm{SIR} = \frac{(9k+11)-(6k+10)}{6k+10} \times 100\% = \frac{3k+1}{6k+10} \times 100\%$$

When k=0, SIR is 10.0%; When k=1, SIR is 25.0%; When k=2, SIR is 31.8%; For the large motion vector, the SIR is close to 50%.

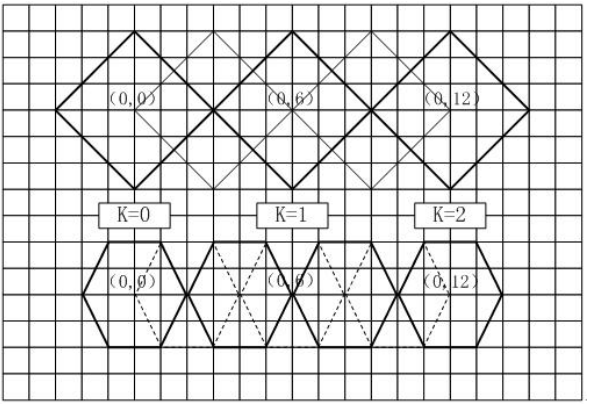


Fig. 5. Assume that the motion vectors are in the horizontal direction with values of (0, 6k); k = 1; 2; 3 . . . : It is easy to see that to find these motion vectors, the number of search points for the LDSPS algorithm follows $6k+10$, whereas the number for HEXBS is $9k+11$.

5 Experimental Results for Comparison

The experimental condition is as follows. The distortion measurement of mean absolute difference (MAD) is used, block size:8×8, and search window size of ± 15. Seven standard

video sequences“Mother”,“Foreman”,“Flower”,“Basket”, “Mobile” (352×288),“Boat”and “Piano” (720×576) were used, which vary in motion content as well as frame size. Average MAD values and average search point numbers are summarized in Tables I and II for different algorithms, including FS, NTSS, FSS, DS, HEXBS and our proposed LDSPS, respectively. Note that only the search region inside the image boundary is considered consistently for all the fast algorithms tested to make a fair comparison.

We can see that the proposed LDSPS consumes the smallest number of search points with just a slight increase in MAD compared with other fast algorithms. The number of search points used by the LDSPS method is substantially smaller than that by NTSS, FSS, DS, or HEXBS, nearly half the number of NTSS. Here, we mainly compare the HEXBS with the proposed LDSPS algorithm in terms of number of search point as well as MAD. According to Tables 1 and 2, Table 3 particularly tabulates the average SIR and average MAD increase in percentage of the proposed LDSPS over HEXBS. For “Mother” sequence with motion vectors limited within a small region around, our proposed LDSPS achieves 11.14% speed improvement over HEXBS. For “Foreman” sequence with medium motion, the average SIR of LDSPS over HEXBS is 11.26%. For “Mobile” and “Boat” which contain large motion, as predicted in theory, our LDSPS has obtained higher speed improvement over HEXBS, here more than 22%. The large the motion in a video sequence, the large the speed improvement rate of LDSPS over HEXBS or the other fast algorithms will be. On the other hand, the change in MAD of LDSPS compared with HEXBS is trivial, some video sequence less than 1.04% or smaller of MAD increase, while the other video sequences’ MAD is decrease as high as 1.30%.

Table 1. Average MAD per pixel for different methods and different video sequence(proposed methods are highlighted)

	Mother	Foreman	Basket	Flower	Mobile	Boat	Piano
FS	1.086	5.884	4.963	5.963	11.656	8.478	2.257
NTSS	1.093	5.900	4.999	6.075	11.706	9.076	2.284
FSS	1.093	5.917	4.985	6.162	11.871	10.635	2.313
DS	1.091	5.920	4.980	6.104	11.734	9.390	2.310
HEXBS	1.096	5.974	4.993	6.228	11.965	9.678	2.312
LDSPS	**1.095**	**5.996**	**5.003**	**6.293**	**11.957**	**9.552**	**2.295**

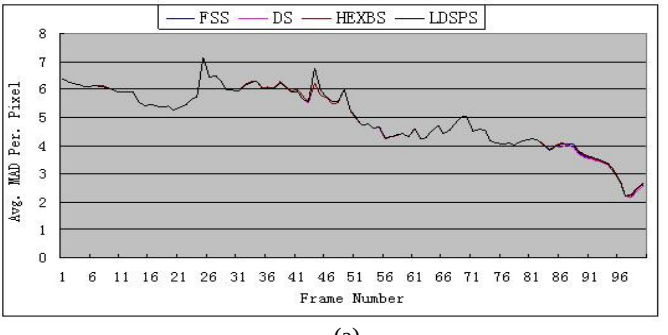
Table 2. Average number of search points per block with respect to different methods and different video sequence(proposed methods are highlighted). Note that the search point number per block for the FS is Fixed as 961 for all video sequences.

	Mother	Foreman	Basket	Flower	Mobile	Boat	Piano
FS	961	961	961	961	961	961	961
NTSS	18.29	18.34	20.61	24.20	22.20	34.93	29.81
FSS	17.52	17.56	19.24	20.55	20.11	23.94	22.65
DS	14.08	13.99	16.95	18.55	17.51	34.76	22.00
HEXBS	11.38	11.46	13.39	14.58	14.19	22.53	16.32
LDSPS	**10.24**	**10.30**	**11.61**	**12.83**	**11.62**	**18.34**	**15.00**

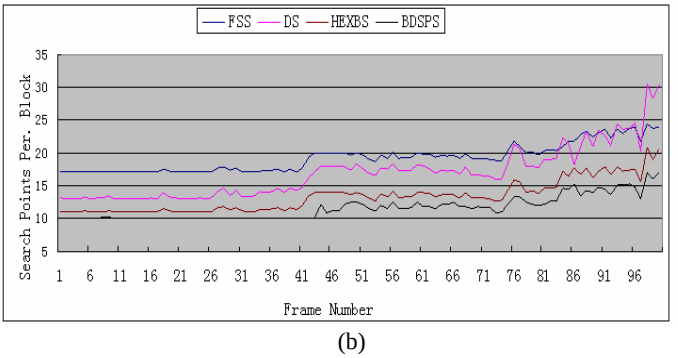
Table 3. Average SIR and Average MAD increase in percentage of LDSPS over HEXBS

	Mother	Foreman	Basket	Flower	Mobile	Boat	Piano
Avg.SIR(%)	11.14	11.26	15.29	13.65	22.11	22.83	8.83
Avg.MAD increase(%)	-0.09	0.37	0.20	1.04	-0.07	-1.30	-0.74

Fig. 6(a) and (b) plot a frame-by-frame comparison of MAD and search point number per block respectively for the different algorithms applied to "Basket" sequence. Fig. 6(a) shows the similar MAD performance for all the methods tested, while Fig. 6(b) clearly manifests the substantial superiority of the proposed LDSPS to the other methods in terms of number of search points used. From Fig. 6(b), we can also see that the curve of the proposed LDSPS fluctuates much less violently than that of HEXBS with respect to the number of search points. Apparently, all the experimental results substantially justify the fastest performance of the proposed LDSPS algorithm as compared with the other popular fast algorithms.



(a)



(b)

Fig. 6. Performance comparison of FSS, DS, HEXBS and the proposed LDSPS for "Basket" sequence in terms of: (a) average MAD per pixel and (b) the average number of search points per block

6 Conclusion

A novel fast algorithm using a large diamond and small pentagon search patterns in block motion estimation is proposed in this paper, which demonstrates significant speedup gain over the HEXBS and other fast search methods while maintaining similar distortion performance. The proposed LDSPS consistently has a faster search performance than DS, regardless of no-, small-, medium-, or large-motion. Strikingly, for large motion image sequences, the new method may use much fewer search points than the HEXBS algorithm. Theoretical analysis shows that a speed improvement of up to 50% over the HEXBS algorithm can be obtained for locating some motion vectors in certain scenarios. In other words, the new algorithm may find any motion vector in motion field with fewer search points than the HEXBS algorithm. Generally speaking, the large the motion vector, the more significant the speedup gain for the new method will be. The experimental results have verified the statement, which have convincingly demonstrated the superiority of the proposed LDSPS to the other fast methods in terms of using the smallest number of search points with a very small penalty of marginal degradation in distortion. Note that based on the LDSPS algorithm and combined the prediction of search beginning point with the search stop criterion ahead of time, motion estimation algorithm may be further speeded up.

References

1. T. Koga, K. Iinuma, A. Hirano, Y. Iijima, and T. Ishiguro.: Motion compensated interframe coding for video conferencing. In Proc. Nat. Telecommun. Conf., New Orleans, LA, Nov. 29–Dec. 3 1981, pp. G5.3.1–5.3.5.
2. R. Li, B. Zeng, and M. L. Liou.: A new three-step search algorithm for block motion estimation. IEEE Trans. Circuits Syst. Video Technol., vol. 4, pp. 438–442, Aug. 1994.
3. L. M. Po and W. C. Ma.: A novel four-step search algorithm for fast lock motion estimation," IEEE Trans. Circuits Syst. Video Technol., vol. 6, pp. 313–317, June 1996.
4. L. K. Liu and E. Feig.: A block-based gradient descent search algorithm for block motion estimation in video coding. IEEE Trans. Circuits Syst. Video Technol., vol. 6, pp. 419–423, Aug. 1996.
5. S. Zhu and K.-K. Ma.: A new diamond search algorithm for fast blockmatching motion estimation. IEEE Trans. Image Processing, vol. 9, pp.287–290, Feb. 2000.
6. J. Y. Tham, S. Ranganath, M. Ranganath, and A. A. Kassim.: A novel unrestricted center-biased diamond search algorithm for block motion estimation. IEEE Trans. Circuits Syst. Video Technol., vol. 8, pp.369–377, Aug. 1998.
7. Ce Zhu, Xiao Lin, and Lap-Pui Chau.: Hexagon-Based Search Pattern for Fast Block Motion Estimation. IEEE Trans. Circuits Syst. Video Technol., vol. 12, pp. 349-355, May 2002.
8. Peng Yang, Yuwen He, Shi Qiang Yang.: An Unsymmetrical-Cross Multi-Resolution Motion Search Algorithm For Mpeg4-Avcm. 264 Coding. IEEE International Conference on Multimedia and Expo (ICME), pp.531-534, 2004.
9. Yang Peng, Wu Hua, Yang Shiqiang.: Fast motion estimation algorithm for H.264. Journal of Tsinghua University(Science and Technology) ,2005, Vo l. 45, No. 4.

Shot Boundary Detection Algorithm in Compressed Domain Based on Adaboost and Fuzzy Theory

Zhi-Cheng Zhao and An-Ni Cai

School of Telecommunication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
Zhao.zc@gmail.com, Annicai@bupt.edu.cn

Abstract. A shot boundary detection algorithm based on fuzzy theory and Adaboost is proposed in this paper. According to changes of color and camera motion, videos are classified into six types. By using features in compress domain such as DCT coefficients, the type of the MB, HSV color histogram difference, camera motion difference and so on, videos are segmented into three classes, that is, cut shot, gradual shot and non-change. The results of experiment have shown that this algorithm is robust for camera motion and walk-in of large objects in videos, and have better precision of shot boundary detection compared with the classic double-threshold method and the method of presented by Kuo *et al*.. There is no problem of threshold selection in our algorithm but it exists in most of other algorithms.

1 Introduction

With the development of the multimedia and internet technologies, content-based video retrieval (CBVR) has been paid more and more attention, and has become the focus of information retrieval field [1]. In many video retrieval algorithms shot is the basic video cell for operation, which means that before depositing and searching, video has to be segmented into shots. Generally, by detecting the boundaries among shots one can confirm the transited positions of shots [2].

The shot is usually categorized into two types: cut shot and gradual shot. Cut shot occurs between two frames of a video sequence, but gradual shot is a slow process from a shot to transit to another. Currently, a majority of videos has been deposited in compressed format, so the shot boundary detection in compressed domain becomes the focus of research.

Now, algorithms presented in the literature for video segmentation in compressed domain are based on DCT coefficients, DC image and MB types, DCT coefficients and motion vectors [3]. Most of these algorithms involve the problem of threshold selection, which directly affects the precision of shot boundary detection. Recently, several researchers proposed such methods as double-threshold algorithm [4], adaptive threshold selection algorithm [5] to improve detection result. However, since features such as camera motion have not been used in these algorithms, the problems of motion sensitive remain unresolved and unable to adapt to various videos.

In this paper, fuzzy theory and Adaboost algorithm [6], [7] are applied to shot boundary detection in compressed domain. Firstly, on the basis of fuzzy theory, we

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 617 – 626, 2006.

use color and motion features which are sensitive to human vision to roughly classify videos into six types, which include color change greatly and motion rapidly, color change slightly and motion slowly. Secondly, according to Adaboost's principle, video frames are classified as cut frame, gradual frame and non-change frame, and then, we use features such as HSV color histogram, camera motion and so on to accomplish cascade training aiming at six different types of videos. By doing this, six types of strong classifiers could be obtained.

Finally, the strong classifiers are adaptively selected according to features of video materials to perform test for shot boundary detection. Moreover, we analyze the roles of various features in Adaboost algorithm, and propose several methods to improve the performance of Adaboost. Since we have selected the camera motion as one of training features, experiments proved that this algorithm is robust for camera motion and walk-in of large objects, and is adaptable to various kinds of videos.

The remainder of this paper is organized as follows. In section 2, the basic classification theory of Adaboost is briefly described. Section 3 presents a novel algorithm of shot boundary detection based on fuzzy theory and Adaboost. Section 4 discusses resampling technology and the improvement of Adaboost. Section 5 is the experiment results, and finally in section 6, the conclusion is made.

2 The Adaboost Algorithm

The Adaboost algorithm is briefly described as follows:

• Given N training samples $\{x_1, y_1\},\ldots,\{x_n, y_n\}$, where $y_i = \{0,1\}$ for negative and positive samples respectively. Each sample has K simple features $f_j(\cdot)(1 \le j \le K)$, and each feature corresponds with a simple two-class classifier:

$$h_j(x) = \begin{cases} 1 & p_j f_j(x) < p_j \theta_j \quad (1 \le j \le K) \\ 0 & otherwise \end{cases} \tag{1}$$

where θ_j is threshold and p_j is shift, $p_j = \{\pm 1\}$ denotes the direction of inequality.

• Initialize weights $w_{1,i} = 1/N$ for each $\{x_i, y_i\} \in S$.

• *For* $t = 1,\cdots,T$:

1) Normalize the weights,

$$w_{t,i} = w_{i,j} \Big/ \sum_{j=1}^{N} w_{i,j} \tag{2}$$

2) For each feature j, train a classifier h_j, with the lowest $\varepsilon_j = \sum_i w_{t,i} \mid h_j(x_i) - y_i \mid$.

3) Choose the classifier h_t, with the lowest error ε_t.

4) Update the weights:

$$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i} \tag{3}$$

where $e_i = 0$, if sample x_i is classified correctly, $e_i = 1$, otherwise, and $\beta_t = \varepsilon_t / (1 - \varepsilon_t)$.

• The final strong classifier is:

$$H(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & otherwise \end{cases} \tag{4}$$

where $\alpha_t = \lg(1/\beta_t)$.

3 Shot Boundary Detection Based on Adaboost

One of the important reasons why Adaboost algorithm can be widely applied to many fields such as face recognition [8] and object detection [9] is that the content of the object in question is usually unitary, and easy to extract a plenty of features needed by Adaboost training. However, as for videos, due to its coverall contents, various motions and changeful video editing means, Adaboost algorithm is difficult to be utilized there. Thus, three problems have to be first considered:

1) Feature extraction, that is how to obtain enough features to meet the training?
2) Training set formation, which means how to reduce distributive dispersancy and confusability of training sample space?
3) Fast training realization, which is how to efficiently decrease training time and improve the efficiency of the algorithm?

3.1 Features Extraction

Features selection always determines the performance of Adaboost classifier. We extract features in compressed domain, and make up training and testing sample sets:

(1) Intra-coded MB rate: For I frame, all MBs are Intra coded; for B and P frames, this parameter varies from zero to the total number of MBs in whole frame. If P frame mostly comprises Intra MBs, it is in high possibility that shot transition occurs in the current frame. a,b,c,d in Fig.1(a) show such cases.

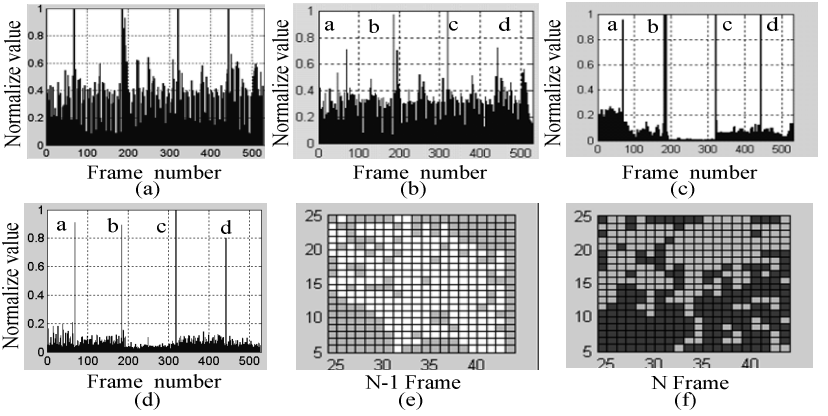


Fig. 1. Several features of the video: (a) the rate of Intra-coded MB of P frame, (b) shot change rate, (c) standard deviation of camera motion, (d) HSV color histogram difference, (e) and (f) distributions of MB type consecutive frames

(2) Shot change rate: It is calculated by backward and forward predicted MBs. By using a video whose length of GOP is 12 and I:B:P is 1:3:8, we can illustrate this feature. Forward reference rate (FRR) is defined as: $FRR = N_f / N$, where N_f is the number of forward predicted MBs and N is the total number of MBs. Backward reference rate (BRR) is defined as: $BRR = N_b / N$, where N_b is the number of backward predicted MBs in current frame. The possible positions of shot change in one GOP are six types, as shown in Fig.2 by character "|".

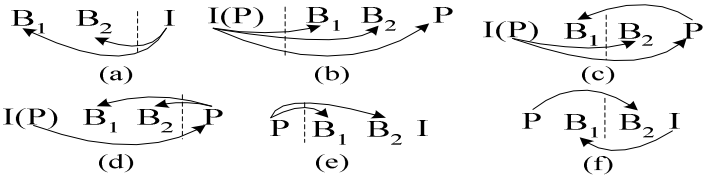


Fig. 2. The possible positions of shot change in one GOP

For example, in Fig.2 (a), if shot change occurs in I frame, the BRRs of the preceding two B frames are small; in Fig.2(c), if shot change takes place in B_2 frame, the FRR of B_2 frame and P frame and the BRR of B_1 frame are small. So the shot change rate shown in Fig.1 (b) could be defined as follows.

$$P_c = 1 - \frac{RR_1^2 + RR_2^2 + \cdots + RR_N^2}{RR_1 + RR_2 + \cdots + RR_N} \tag{5}$$

where, RR denotes FRR or BRR of one frame. The Bigger P_C, the higher the possibility of the shot change is, and P_C=1, if RR_i=0.

(3) Rate of bidirectional predicted MB: For I/P frame, this parameter is zero, and for B frame, the smaller the rate, the larger the shot change occurring possibility is.

(4) Luminance histogram difference between previous and current frames: The bigger the difference, the larger the shot change occurring possibility is.

(5) RGB color histogram difference between previous and current frames: In RGB color space, R,G,B are quantized into 4:4:4 respectively. The bigger the difference, the larger the shot change occurring possibility is.

(6) HSV color histogram difference between previous and current frames: In HSV color space, H,S,V are quantized into 8:3:3 respectively. The bigger the difference, the larger the shot change occurring possibility is, as shown by a,b,c,d in Fig.1 (c).

(7) Differences of DCT coefficients in Y,U,V spaces: The bigger the difference, the larger the shot change occurring possibility is.

(8) Rate of motion MB: That is defined as the radio between the number of MBs with motion vectors not equal to zero and the total number of MBs in one frame. The bigger the rate, the larger the shot change occurring possibility is.

(9) Standard deviation of position after global motion compensation: Suppose that (e_i^x, e_i^y) are the coordinate of a pixel after global motion compensation, i.e.,

$$e_i^x = x_i - p_1(x_i + q_3) \tag{6}$$

$$e_i^y = y_i - p_1(y_i + q_4) \tag{7}$$

where (x_i, y_i) is the coordinates of the pixel, p_1,q_3,q_4 are three parameters for zoom, pan, and tilt of camera [10], respectively. Then we can calculate the standard deviation of position:

$$diff_{ei} = \sqrt{diff_{e_{xi}}^2 + diff_{e_{yi}}^2} \tag{8}$$

where $diff_{e_{xi}}^2$, $diff_{e_{yi}}^2$ are variances of e_i^x and e_i^y, respectively. Example are shown in Fig.1.(d) with high values at a,b,c,d.

(10) Difference of dominant color: DC images are extracted firstly, and subsampling is applied to obtain dominant color for each frame. The bigger the difference of adjacent frames, the larger the shot change occurring possibility is.

(11) Spatial distribution difference of the MB types: The alteration number of MB type of adjacent frames see Fig.1 (e) and (f) is calculated, the bigger the difference of adjacent frames, the larger the shot change occurring possibility is.

However, the number of features is quite insufficient if we only use the above 11 kinds of features to train Adaboost classifier. Further processing is needed. We divide each frame into 6, 9, 12, 16, 18 or 24 sub-blocks, and then select a set of sub-blocks with certain patterns to make up new features. Finally, in our experiment, 120 features are obtained for each frame and formed to be one feature vector after normalization. Fig.3. shows an example of a video frame divided into 9 blocks, and the black blocks are selected to extract new features of the frame.

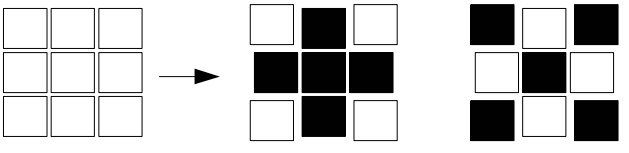


Fig. 3. An example of 9 sub-blocks, and new features are extracted from the black blocks

3.2 Fuzzy Classification for Videos and Training Set Construction

A feature set is comprised by feature vectors which are extracted in turn from every frame of a video. A number of feature sets are stochastically interlaced with each other frame by frame to construct a training sample set. Because the stochastic interlacement partly eliminates temporal correlation of video frames, the training efficiency is elevated to some extent. However, the variety of video materials still limits the effect of interlacement. Therefore, in order to improve the adaptability of shot segmentation algorithm to various videos, we extract two features of color and motion, which are most sensitive to the human vision, to first carry out a fuzzy classification on videos.

Suppose that there are M various video segments, $V_1, V_2, \cdots V_M$, and their lengths respectively are $L_1, L_2, \cdots L_M$, we apply features (6) and (9) which are clarified in

section.3.2 to calculate means $\mu_1, \mu_2, \cdots \mu_M$ and standard deviation $\sigma_1, \sigma_2, \cdots \sigma_M$ of each feature which will then be used to respectively recalculate their means as μ and σ for the M videos. Then we design a Gaussian membership function to analyze features:

$$f(x) = 1 - \exp(-(\frac{x-\mu}{4\sigma})^2) \tag{9}$$

The analysis results of Gaussian membership function are shown in Fig.4.

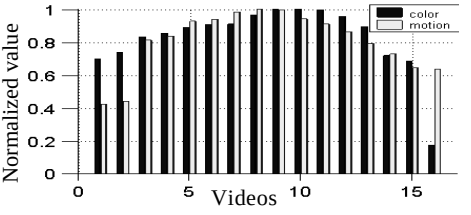


Fig. 4. Using Gaussian membership function to analyze color and motion features

According to the statistics of 24 various videos and the membership functions, we classify them into 6 types. The fuzzy classification is shown in Table.1.

Table 1. Fuzzy classification

Color $f_c(x)$	Motion $f_m(x)$	Classification
$> T_c$	$> T_{m1}$	Great change of color, fast motion
$> T_c$	$> T_{m2}$ & $\leq T_{m1}$	Great change of color, moderate motion
$> T_c$	$\leq T_{m2}$	Great change of color, slow motion
$\leq T_c$	$> T_{m1}$	Slight change of color, fast motion
$\leq T_c$	$> T_{m2}$ & $\leq T_{m1}$	Slight change of color, moderate motion
$\leq T_c$	$\leq T_{m2}$	Slight change of color, slow motion

In the experiment, we select $T_c = 0.4$, $T_{m1} = 0.8$, $T_{m2} = 0.4$. The result of fuzzy classification is quite consistent with human subjective perception. It is shown that fuzzy principle made up of color and motion features could reflect some of the characteristics of the video. And more importantly, fuzzy classification reduces dispersancy and confusability of training sample space and increases efficiency.

3.3 Fast Adaboost Algorithm

Adaboost is a process of machine learning which calculates strong classifier through iterative algorithm with slow convergence. Thus some improvement has to be made in Adaboost algorithm in order to realize fast training. We adopt the following methods:

(1) Eliminating shift parameter p_j in Eq.(1) and making features monotonic. We change all the normalized features into monotonous function and as a result, the

bigger the feature's value, the higher shot change possibility occurs. This method decreases half of the calculation.

(2) Converting three-class classifier into two serial two-class classifiers which will be used to classify cut shot and gradual shot respectively. We find that the training time is reduced to one percentage of the original training time.

(3) Adopting Quick-sort algorithm to eliminating the repeated feature values.

(4) Setting the starting point of threshold which will not search from the smallest value but from $T_s = \overline{\mu} + c \cdot \overline{\sigma}$, where, $\overline{\mu}$ is means of the feature, $\overline{\sigma}$ is its standard deviation, and C is a constant. The approach is proved to decrease the time of convergence.

(5) Using a look-up table of the error rate. We first calculate error rates of corresponding thresholds for each feature and save them in a table. Because the searching by table while training avoids the recalculation for each feature, training time is decreased significantly.

4 Re-sampling and Weight Adjustment

The result of training depends on the coverage of training samples to some extent. In order to solve the contradiction between finity of training sample and diversity of actual video, we use re-sampling technology to accomplish cascade training by stochastically dividing a training set into three sub-sets. Suppose a false positive sample rate at each stage is $p_i \le 8\%(i = 1,2,3)$, after the three training stages shown in Fig.5, the total false positive sample rate will decease to $p = p_1 p_2 p_3 \le 0.05\%$.

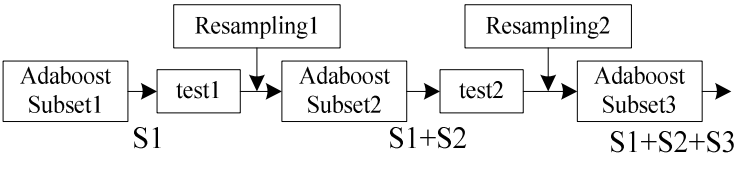


Fig. 5. Flowchart of re-sampling and cascade training

In the experiment, we train classifiers respectively for cut shot and gradual shot. Firstly, a two-class classifier is trained to classify videos into cut shot and non-change shot, where the cut frame is regarded as the positive sample, while the non-change frame as the negative. Secondly, the same method is applied to training for gradual shot. After the first training stage, a test is performed and the false samples are added into the second training stage, where the training and testing follow. Finally, a strong classifier is formed after three stages.

We test 5 videos with this method, and count those features which are corresponding with each weak classifier expressed in Eq.(4). Fig.6 shows the roles that various features played in shot boundary detection. The leading 8 features respectively are: (6), (10), (7), (5), (2), (4), (9) and (11) in section 3.1.

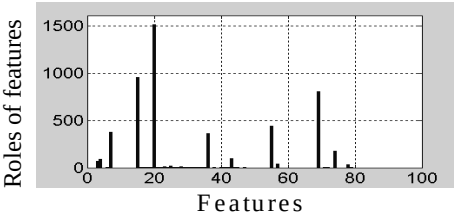


Fig. 6. Roles of features in shot boundary detection

Based on the result above, we adjust the role factor α_t in Adaboost algorithm. The contribution of each feature in strong classifier as shown in Fig.6 is normalized and used as weight factor ω_i to produce a new role factor $\alpha_t' = \alpha_t \omega_i$, $t = 1, \cdots, T$, $i = 1, \cdots, 11$. Thus a new strong classifier is changed into:

$$H'(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t' h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t' \\ 0 & otherwise \end{cases} \tag{10}$$

5 Experiment Results

According to Fig.7, the final test is done. Considering some editing characters of videos, we perform a post-processing on the classification results of each video, and the principles is: if two successive frames are both detected as cut frames, the last frame is selected; if the number of successive gradual shot detected frames is less than 5, these frames will then be deleted from the gradual shot set; if there exist no more than two non-change frames among successive gradual frames, they will be changed into gradual frames.

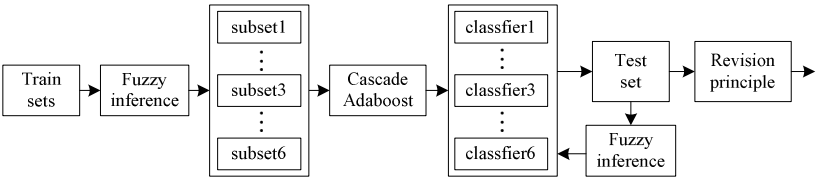


Fig. 7. Flowchart of shot boundary detection

Recall rate and precision rate are often taken as the evaluation measures in shot boundary detection, which are defined as follows:

Recall=correct detection / (correct detection + missed detection).

Precision= correct detection / (correct detection + false detection).

Moreover, Cover is another evaluation measure for gradual shot detection, which is defined as the overlapped part between the detected and actual gradual shot durations in percentage.

In the experiment, we adopt various MPEG-2 video materials such as Ads, sports, scenery and cartoon, which include diverse gradual shot modes, many camera motion types and walk-in of large objects. The result is shown in Table.2.

Table 2. Results of shot boundary detection based on Adaboost and fuzzy theory. The character "C" denotes cut shot, and character "G" denotes gradual shot.

Video	Frame	C	G	Missed		False		Results %				
				C	G	C	G	Precision		Recall		Cover
Ads	7598	292	4	8	1	6	0	C	G	C	G	G
Scenery	7181	37	67	2	6	1	2					
Cartoon	1311	33	1	2	0	1	0	96.1	89	97	94.3	91
Sports	6006	123	6	6	0	6	2					

Using the same video material, we compared our algorithm with the classic double -threshold method and the method presented by Kuo *et al*. [11].

Table 3. Results of double-threshold algorithm and the method of presented by Kuo *et al.*. The character "C" denotes cut shot, and character "G" denotes gradual shot.

Method	Shot		Correct		Missed		False		Precision%		Recall %	
	C	G	C	G	C	G	C	G	C	G	C	G
Double-threshold	251	56	212	35	16	13	23	8	90.2	81.4	93	73
Kuo	116	35	78	22	11	7	17	6	82.1	78.5	87.6	75.8

The experiments show that good result can be obtained in slowly-changed videos such as scenery when using the method of classic double-threshold based on color histogram and the method presented by Kuo *et al*., however, for videos with changeful motions and complicated forms, for instance, sports and ads, the results is of these algorithms are not satisfactory. It indicates that their applications are limited, and moreover, the problem of threshold selection exists when only one or two features are used to detect shot boundary. Nevertheless, compared with these two, our algorithm is robust for camera motion and walk-in of large objects in videos, and obviously has a better precision and recall in shot boundary detection.

6 Conclusion

In this paper we use fuzzy theory to classify video material and adopt Adaboost algorithm to accomplish shot boundary detection. The results of experiment have shown that this algorithm is robust for camera motion and walk-in of large objects in videos, and have better precision of shot boundary detection compared with the classic double-threshold method and the method presented by Kuo *et al*.. Our algorithm has no problem of threshold selection which exists in most of other algorithms. More significantly, based on Adaboost algorithm, we efficiently fuse the features extracted from

compressed domain, and analyze the roles of features played in shot segmentation, and ameliorate role factor of Adaboost according to the contributions of features. In the end, we use re-sampling and cascade training technologies to improve classification result.

It's necessary to point out that the classification result in Adaboost algorithm is affected by human subjective judgement. Thus, how to set up an appropriate visual model to improve reliability of this algorithm is our further task.

References

1. Over, P., Ianeva, T.: TRECVID 2005 - An Overview. In: Proceedings of TRECVID 2005 Workshop. March (2006)
2. Mezaris, V., Kompatsiaris, I.: Real-Time Compressed Domain Spatiotemporal Segmentation and Ontologies for Video Indexing and Retrieval. IEEE Trans. on Circuits and Systems for Video Technology. Vol.14 (2004) 606–621
3. Koprinska, I., Carrato, S.: Tepmoral Video Segmentation: A survey. EURASIP Sig.Proc.: Image Communication. Jan.5 (2001) 477–500
4. Zhang, H.J., Kankanhalli, A.: Automatic Partitioning of Full-motion Video. Multimedia System(1).(1993) 10–28
5. Boccignone, G., Santo, M.D.: Automated Threshold Selection for Detection of Dissolves in MPEG Videos.IEEE International Conference on Multimedia and Expo(ICME).July.1(2000)
6. Freund, Y., Schapire, R. E.: A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Science (55).(1997) 119–139
7. Schapire, R.E.: The Boosting Approach to Machine Learning an Overview. MSRI. Workshop on Nonlinear Estimation and Classification (2002)
8. Yang, P., Shan, S.G., Gao, W.: Face Recognition Using Ada-Boosted Gabor Features Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR'04). (2004)
9. Paul, V., Michael, J.: Robust Real-time Object Detection. Second International Workshop on Statistical and Computational Theories of Vision – Modeling, Learning, Computing, and Sampling Vancouver, Canada, July.12 (2001)
10. Tan, Y.P., Saur, D.D., Kulkarni, S.R.: Rapid Estimation of Camera Motion from Compressed Video with Application to Video Annotation [J]. IEEE Trans on Circuits and Systems for Video Technology. Jan.1 (2000) 133–146
11. Kuo, T.C.T., Lin, Y.B.: Efficient Shot Change Detection on Compressed Video Data. IEEE Trans. On Circuit and Systems for Video Technology May.5 (1996) 101-108

A Novel Unified SPM-ICA-PCA Method for Detecting Epileptic Activities in Resting-State fMRI*

Qiyi Song[1], Feng Yin[2], Huafu Chen[1], Yi Zhang[1], Qiaoli Hu[1], and Dezhong Yao[1],

[1] Center of Neuroinformatics, School of Applied Mathmatics,
University of Electronic Science and Technology of China,
Chengdu, 610054, PR China
uestcsqy@163.com, {chenhf,Dyao}@uestc.edu.cn
[2] Department of Mathematics,
Sichuan University of Science and Engineering,
Zigong, 643000, PR China

Abstract. In this paper, it is reported that the method and primary application of a novel noninvasive technique, resting functional magnetic resonance imaging (fMRI) with unified statistical parameter mapping (SPM) independent component analysis (ICA), and principal component analysis(PCA), for localizing interictal epileptic activities of glioma foci. SPM is based on the general linear model (GLM). ICA combined PCA was firstly applied to fMRI datasets to disclose independent components, which is specified as the equivalent stimulus response patterns in the design matrix of a GLM. Then, parameters were estimated and regionally-specific statistical inferences were made about activations in the usual way. The validity is tested by simulation experiment. Finally, the fMRI data of two glioma patients is analyzed, whose results are consisting with the clinical estimate.

1 Introduction

SPM, based on the GLM, is a powerful tool for the analysis of functional mapping experiments [1]. To measure the magnitude of the blood-oxygenation-level-dependent (BOLD) signal that is task-specific, neuroimaging data at each voxel is modeled as a linear combination of explanatory variables plus a residual error term [2]. SPM creates images of a statistic reflecting 'significance'. These SPMs is interpreted as spatially extended statistical processes that behave according to the theory of Gaussian fields [3]. This enables the statistical characterization of regionally specific responses (e.g., using t tests or F tests). This technique makes it possible to test multiple factors that may contribute to the signal changes in neuroimaging data [4].

SPM is a model-driven and depends on some hypotheses about the data. These hypotheses are embodied in the design matrix of the GLM. The application of the GLM precedes under two assumptions: normal distribution and independence of the error

* Supported by NSFC#90208003, #30200059, 30470510, TRAPOYT, the 973 Project No. 2003CB716100, Doctor training Fund of MOE, PRC, Fok Ying Tong Education Foundation (91041).

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 627 – 636, 2006.

terms. Furthermore, SPM is a univariate approach, because it characterizes each voxel separately and performs voxel-wise statistical analyses in parallel. In neuroimaging, several factors can change the observed data. [4].

SPM offers several options to model evoked changes in signal intensities [5], including a canonical hemodynamic response function (HRF), which can be supplemented with various derivatives. These basis functions are used to create regressors in the design matrix. The ensuing GLM is a convolution model that depends on knowing the form of the HRF and the stimulus functions. However, assumptions about the HRF are not always valid and different stimuli may elicit different kinds of hemodynamic responses [6]. If the assumed forms of the HRF, or the stimulus functions it is convolved with, are incorrect or incomplete, this may result in biased estimates of the true response [7]. Rest is the state that in the brain when it is not engaged in specific goal-oriented task. So SPM is not fit for detecting activation sources when the data is acquired in the resting-state.

Complementary methods, driven by the data, do not make any assumption about the causes of responses or the form of the HRF. They have been applied to functional mapping experiments in the context of principal component analysis (PCA) [8], Blind source separation (BSS) [9], and clustering analysis [7]. These methods emphasize the intrinsic structure of the data. An essential difference between these data-driven approaches and SPM is that they are multivariate approaches accounting for interactions among different regions or voxels [4].

Among multivariate approaches, ICA has attracted attention recently and represents a promising approach to characterizing evoked responses [10]. ICA is capable of extracting multiple sources such as task-related components, cardiac and respiratory effects, subject movements, and noise. Two ways of the application are spatial-ICA [11] and temporal-ICA.

Nevertheless, there are limitations to purely data-driven approaches. It is difficult to put these approaches into a statistical framework that allows one to test the activations against a desired hypothesis. Furthermore, the ICA algorithm has an inevitable problem: how to select the Independent Components (ICs) that describe “most relevant” independent directions within the training dataset, a sorting method for ICA is required. This question is still open in the literature. In addition, it lacks the ability to assess the local or regionally-specific nature of brain responses. Recently, Dewen Hu, propose the unified SPM-ICA method, to analyze block stimulation fMRI data, suggest that the method has a higher sensitivity than SPM analysis [4]. This approach successively combines independent components to construct task-related components and then turns to a fully hypothesis-driven approach.

But epilepsy is characterized by transient behavioral and electro-physiologic disturbances, which may or may not be associated with detectable structural lesions. In the localization of epileptic foci by resting fMRI, the timing of the stimuli (endogenous epileptic activities) throughout the imaging series is unclear, so the method is proposed by Dewen Hu can not apply to the epileptic localization.

PCA utilizing the orthogonality between signal and noise can separate the fMRI data into signal subspace and noise subspace. PCA results in a natural ordering of the vectors according to their variance, eigenvectors with highest variance are placed front and those with low variance are placed last. The resulting principal components which construct the signal subspace focalize the mass signal elicited by physiological

or non-physiological factors, thus it can be regarded as the pattern of activation signal. Furthermore, ICA has the performance of extracting multiple sources such as task-related components, cardiac and respiratory effects. For the resting-fMRI are often local area activity and the signal noise ratio is higher such as many kinds of epilepsy demonstrate intense activation in resting-state, so we can select the components derived from ICA which has tightly correlation with the resulting principal component as equivalent stimulus response pattern to model the resonance of equivalent explanatory variables. Furthermore, SPM is a fairly mature framework for neuroimaging data analysis and has been applied successfully in many situations. It would be nice to harness the inferential power of SPM to make inferences about data-led responses in a regionally-specific fashion [4]. With an eye to this issue, we present a united univariate SPM, multivariate temporal ICA and PCA to localize the foci of glioma. In this method, the activation strength formula for index the cortex activity area and the equivalent stimulus response pattern instead of some column vectors of the SPM design matrix. Parameters were estimated and regionally –specific statistical inferences were made about activations in the usual way.

The method is presented in the next section, followed the simulation analysis, the primary application on two glioma patients in section, the final is the conclusion and discussion section.

2 A Novel Unified Method

The fMRI data are composed of functional active signals and noisy background. Let $y_p(t)$ denote the fMRI measured data at voxel p, and then $y_p(t)$ is:

$$y_p(t) = s_p(t) + n_p(t) \tag{1}$$

Where p =1… k is the index of voxel, $s_p(t)$ is functional activities, and $n_p(t)$ are noises of each voxel. The matrix form of Eq. (1) is:

$$y = s + n \tag{2}$$

Where $y \in R_{\text{K*N}}$, $s \in R_{\text{K*N}}$ and $n \in R_{\text{K*N}}$, N stands for the number of temporal sample point, K is the number of voxel, and the subscript p is the index for voxel.

We first segment the fMRI temporal signals into background signals and non-background signals by simple thresholding. The term local area is made up of the calculating voxel and the 4 nearest neighbor voxels, respectively.

2.1 Remain Step of the Novel Algorithm Are Briefly Described as Following:

1. The delay principal component analysis for the temporal signal is performed on global fMRI dataset to decrease noise.
2. Calculating activation strength at every local area according to the following criterion:

$$\text{Activation}(k) = \frac{\sum_{j=1}^{5} \sigma_{j,k}}{\sum_{i=1}^{m} \sigma_i} \times 100\% \tag{3}$$

Where the symbol σ stand for variance, the subscript k is the index for the calculating voxel, and the subscript i is the index for every voxel. Only the activation strength exceed given threshold can be selected as potential activation area.

3. The procedures of conducting local PCA (LPCA) and local ICA (LICA) on each local fMRI data set were performed, respectively. We select the component as the equivalent stimulus response patterns for it has the largest correlate coefficient with the first principal component and also has large correlate coefficient with itself time-course. We apply the FastICA algorithm to the ICA decomposition procedure.

4. For SPM, we used the conventional method to construct the design matrix. A box-car stimulus function convolved with a canonical HRF was included.

5. We applied the equivalent stimulus response patterns instead of some column vectors of the design matrix. Parameters were estimated and regionally-specific statistical inferences were made about activations in the usual way.

3 Simulation and Analysis

3.1 Set the Cortical Activities and Generate Background Noise

The rest state is characterized by an organized baseline level of activity [12] the null dataset representation rest state was acquired from a healthy subject who performed no specific task. The fMRI data are composed of baseline activity, BOLD signal and noisy background. Every fmri slice corresponds to a 64*64 image matrix. A stimulus was periodically applied on for 14s then off 14s and for total 240s. There are three type cortical BOLD activities, the first one as following: 1816 (24, 29), 1369 (25, 22), 2388 (20, 38), 3632 (48, 57), the second one as following: 1754 (26, 28), 3344 (36, 53), 875 (43, 14), 1558 (20, 25), and the third one as following: 1186 (34, 19), 1425 (17, 23), 1557 (21, 50), 2608 (48, 41).The distribution of the twelve sources is shown in Fig1. Following references [13] and [14], we used passion kernel to model the hemodynamic response h (τ) = (μ τ)*e-μ / (τ !), Where μ represents the delay in the response to the stimulus. In this simulation, we set μ at 8s for the first type, 11s for the second type, and 6s for the third type. We modeled the intensity distribution of these cortical activities by using a Gaussian distribution; its full-width at half-maximum was set at 0.2 cm based on the investigation reported in [15]. The maximum of the BOLD activities are no more than 4% of the baseline level activity. The standard deviation of the white noise added to each voxel has approximately 1% of background constant. The BOLD response curves of the 1816 (24, 29) of the cortex is shown in Fig.2.

Fig. 1. The assumed twelve sources in computer simulation

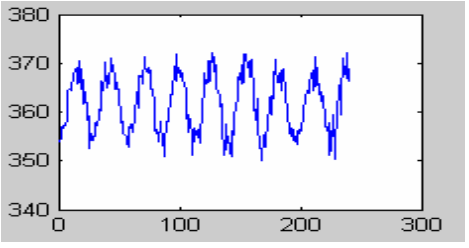


Fig. 2. The time curves of source 1816

3.2 Calculate the Activities Strength Coefficient

In this stimulation, we calculate the activation strengthen coefficient at every local area as formula (3), and the activation strength one-dimension curve is shown in Fig.3. The two-dimension activation map is depicted in Fig.4.

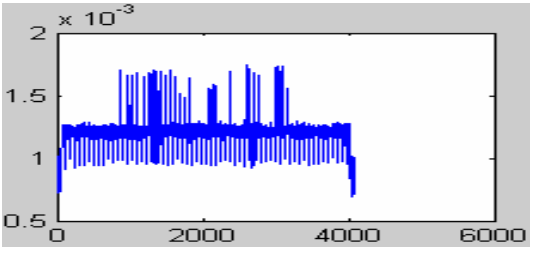


Fig. 3. The activation strength curve by using the novel criterion

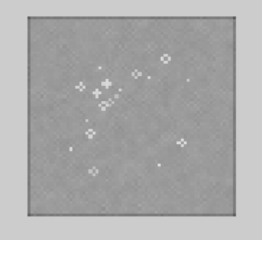


Fig. 4. The activation map obtained by using the novel criterion

3.3 Calculate the LPCA and LICA

In this simulation, these voxel whose activation strength exceed given threshold are considered as potential activation voxel. Several different methods for calculating ICA exist, such as the FastICA, the InfoMAX and the JADE algorithm. These methods differ in the optimized contrast function to achieve de-correlation. The Fast-ICA method is an iterative fixed-point algorithm that was developed by Hyvärinen et al. It is an alternative to gradient-based methods for maximizing non-Gaussianity and shows fast (cubic) convergence. We conducted independent decomposition and principal decomposition on every potential activation area. Select the ICA component it has maximums correlation coefficient with the first principal component as the potential equivalent stimulus time course.

Some components derived from FastICA are shown in the Fig.5. Correlation coefficient curve of the principal component and the selected Independent component is shown in Fig.6. Then we calculate the correlation coefficient of the selected Independent components with the voxel time course, if correlation coefficient exceed the

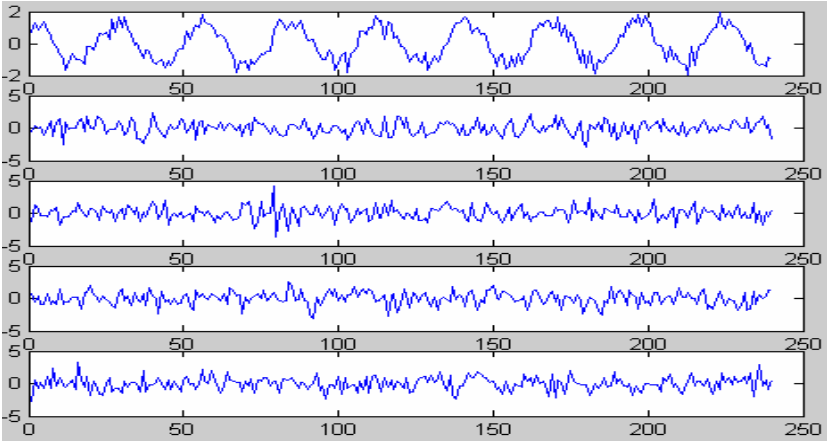


Fig. 5. Some independent components derived from the FastICA decomposition

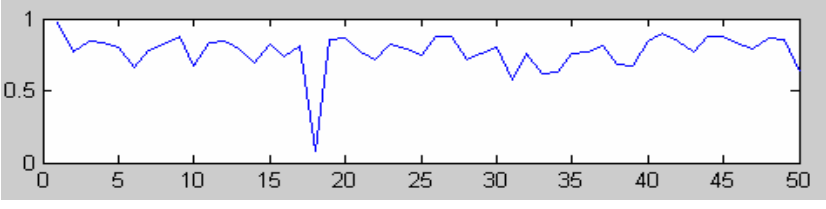


Fig. 6. The correlation coefficient curve of the principal component and the selected equivalent stimulus response patterns at all potential activities areas

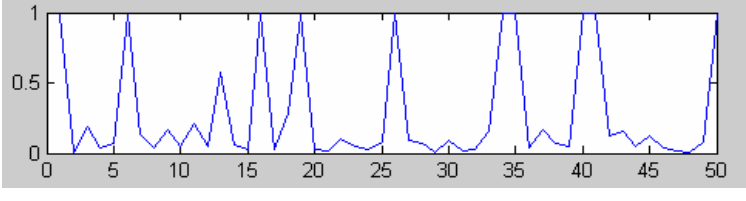


Fig. 7. The correlation coefficient curve of every equivalent stimulus response pattern with itself real time curve, respectively

threshold are consider as the equivalent stimulus time course. The correlation coefficient curves of the equivalent stimulus time course is shown in Fig.7.

3.4 Equivalent Stimulus Response Patterns Instead of the Design Matrix and Analysis by SPM Software

We applied the equivalent stimulus patterns instead of some column vectors of the design matrix, and then did some analysis by SPM software. The result is shown in figure8 and result analysis as follows:

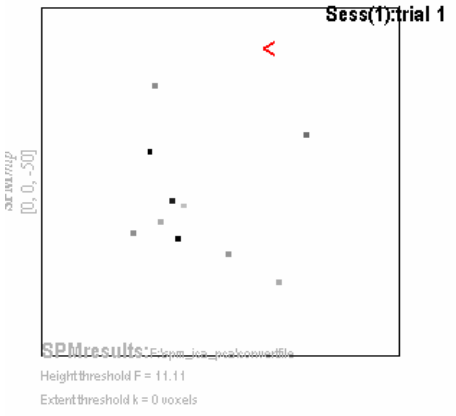


Fig. 8. The SPM-ICA-PCA result map

From the map of activation derived from SPM software, we can find that the voxel 3344 (36, 53) and the voxel 3632 (48, 57) weren't detected. For the two voxels signal noise ratio are lower than the given threshold. The total signal variance is approximately the sum of all components variance; hence we can use the variance percent of local area to index the 'activation significance'. With ICA such an ordering is not obtained automatically. The components derived from ICA often represent independent sources including the heart beat, breathing and head motion artifacts. However, when brain is in the state of rest, the signals are including spontaneous low-frequency BOLD signal and some other noise (e.g., cardiac and respiratory effects). These noises except that less well understood sources such as low frequency drifts and high amplitude oscillations caused by pulse effects that induce localized motion of brain tissue are often difficult to localize into special area. If there contains fluctuation source, the BOLD signal is significantly stronger than noise in local area, that is the activation area with big signal noise ratio. So we can consider the maximal variance component representing the activities source which result in BOLD signal. If local area exists fluctuation source which contribute to functional magnetic resonance (fMRI) signal, PCA result in the first component representation almost activities signal which including BOLD signal. The basic idea is to concatenate the data from several voxels in the spatial dimension, and perform ICA to identify common activation time course. So we can consider the independent component which has the maximal correlation coefficient with the first principal component in identical local area as the fluctuation BOLD signal. By using spatial ICA in local area, we can acquire the temporal pattern, the correlation coefficient curve of the selected component with the time course of given activities voxel and also validate that the selected component can be equivalent stimulus response patterns of BOLD signal. The activation map derived from SPM can also validate the accuracy of this novel algorithm.

4 Application

The fMRI data were provided by Nanjing general hospital of PLA. These whole brain BOLD/EPI images were acquired on a modified 2T Siemens MAGNETOM vision system. Each acquisition consists of 23 contiguous slice (64*64*23, 3.75mm*3.75mm*4.00mm) voxel images. The scan to scan repeat time (RT) is set to 2000ms (Epochs of 200 scans) The total 200 BOLD-fMRI images were acquired while subject was at rest with eyes closed in a period of 400s scan time.

When processing this data, we firstly extracted data by the value larger than 200. By calculating the covariance matrix with 2 second time delay, then conducting SVD

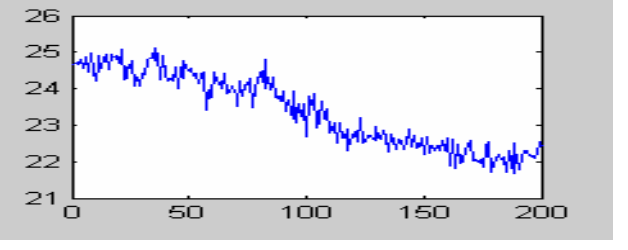


Fig. 9. Equivalent stimulus pattern

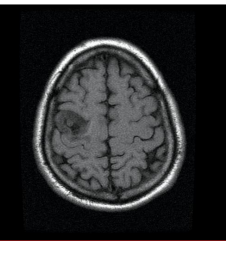


Fig. 10. Anatomy image

The regions of interest derived from SPM, are shown as follows:

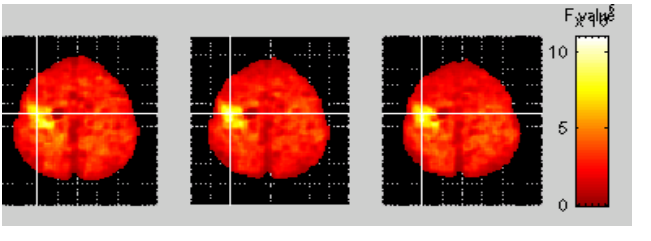


Fig. 11. Depicts the results derived from SPM consist with the foci of anatomy demonstrated.

The other example of glioma epilepsy is shown as following:

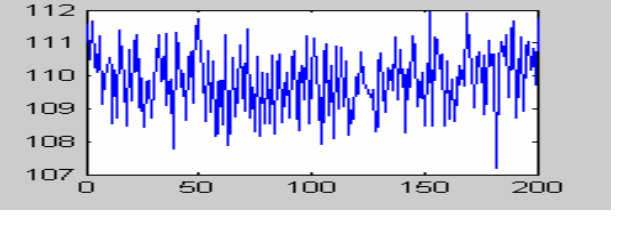


Fig. 12. Equivalent stimulus pattern

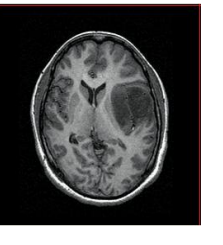


Fig. 13. Anatomy image

on the matrix and choose the first component as principle component. We can get local principal components and local independent component, by conducting LPCA and ICA on corresponding local fMRI dataset, respectively. In this algorithm, ICA was used as implemented by the software package Fast-ICA V.23. We extract the equivalent stimulus pattern by using the proposed algorithm, and it is shown in Fig.9, the anatomy image was shown in Fig. 10.

Epilepsy is characterized by transient behavioral and electro-physiologic disturbances. Glioma is account for 40% of epileptic source. Many kinds of epilepsy, such as glioma, are associated with the structural lesions. Glioma is primary central nervous system tumors that arise from oligodendrocytes, astrocytes or their precursors. Orgnic pathologic changes often involve enhanced variation of metabolism, it lead the BOLD signal involves enhanced variation, so we can detect BOLD activation by using PCA [16] algorithm and seek the equivalent stimulus pattern, and then utilized the SPM (soft) to make statistic analysis. It has shown that the results derived from the application of our algorithm to vivo glioma fMRI data are conformed to those determined by the presurgical assessments.

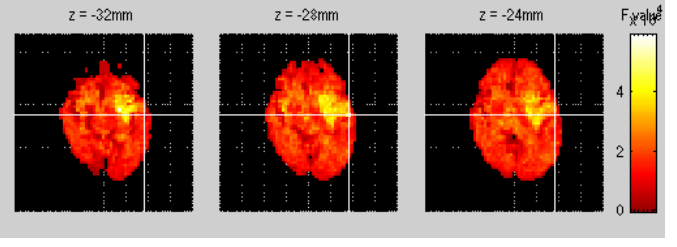


Fig. 14. Depicts the results derived from SPM consist with the foci of anatomy demonstrated

5 Discussion and Conclusion

A novel unified SPM-ICA-PCA algorithm is presented for activation signal detection of resting-state functional magnetic resonance imaging (fMRI). In this algorithm, a new criterion is proposed for calculating the activation strength and a novel equivalent stimulus response pattern is acquired by uniting local principal component analysis with local independent component analysis. By applying the equivalent stimulus pattern into SPM software, the activation sources of resting-state we can detect accurately. Meanwhile, computer stimulation and vivo epilepsy fMRI data analysis demonstrate the potential of this technique to localize foci (e.g., epilepsy) in resting-state fMRI. What the most important is, this paper provide the chance of utilize the novel algorithm to detecting activation source by using SPM soft in rest fMRI especially in the patients with serious brain diseases.

References

1. Friston, K.J., Frith, C.D., Turner, R., Frackowiak, R.S.: Characterizing evoked hemodynamics with fMRI, Vol.2, NeuroImage (1995) 157—165
2. Friston, K.J., Holmes, A.P., Poline, J.B., Grasby, P.J., Williams, S.C., Frackowiak, R.S., Turner, R.: Analysis of fMRI time-series revisited, Vol. 2, NeuroImage (1995) 45—53

3. Adler, R.J.: The geometry of random fields, John Wiley & Sons, Inc., New York (1981)
4. Dewen, H., Lirong, Y., Yadong, L., Zongtan, Zh., Friston, K.J., Changlian, T., Daxing, W.: Unified SPM-ICA for fMRI analysis, Vol.25, Neuroimage (2005) 746-755
5. Della-Maggiore, V., Chau, W., Peres-Neto, P.R., McIntosh, A.R.: An empirical comparison of SPM preprocessing parameters to the analysis of fMRI data, Vol. 17, NeuroImage (2002)19—28
6. Friston, K.J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M.D., Turner, R.: Event-related fMRI: characterizing differential responses, Vol.7, NeuroImage (1998) 30—40
7. Fadili, M.J., Ruan, S., Bloyet, D., Mazoyer, B.: A multistep unsupervised fuzzy clustering analysis of fMRI time series, Vol. 10, Hum. Brain Mapp (2000) 160—178
8. Backfrieder, W.: Quantification of intensity variations in functional MR images using rotated principal components. Vol.41, Phys. Med. Biol (1996) 1425-1438
9. Stone, J.V.: Blind source separation using temporal predictability, Vol.13, Neural Compute. (2001) 1559—1574
10. Calhoun, V.D., Adali, T., Pearlson, G.D., Pekar, J.J.: Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms, Vol. 13, Hum. Brain Mapp (2001) 43—53
11. McKeown, M.J., Makeig, S., Brown, G.G., Jung, T.P., Kindermann, S.S., Bell, A.J., Sejnowski, T.J.: Analysis of fMRI data by blind separation into independent spatial components, Vol. 6, Hum. Brain Mapp (1998) 160—188
12. Fransson, P.: Spontaneous Low-frequency BOLD Signal Fluctuations: An fMRI Investigation of the Resting-state Default Mode of Brain Function Hypothesis, Vol.26, Human Brain Mapping (2005)15-29
13. Friston, K.J., Jezzard P., Turner, R.: Analysis of functional MRI time series, Vol.1, Hum. Brain Mapp (1994) 153—17
14. Quyang, X., Pike, G.B., Evance, A.C.: FMRI of human visual cortex using temporal correlation and spatial coherence analysis, Proc., SMR, 2nd Annual Meeting, San Franciseco (1994) 633
15. Turner, R., Grinvald, A.: Direct visualization of patterns of deoxygenating and deoxygenating in monkey cortical vas-culture during functional brain activation, Proc., SMR, 2nd Annual Meeting, San Franciseco, (1994) 430
16. Zhang Y., Ye M, Lv J C., Tan K K. Convergence analysis of a deterministic discrete time system of Oja's PCA learning algorithm, 16(6), IEEE Trans on Neural Networks (2005), 1318-1328

Design IIR Digital Filters Using Quantum-Behaved Particle Swarm Optimization

Wei Fang, Jun Sun, and Wenbo Xu

Center of Intelligent and High Performance Computing
School of Information Technology, Southern Yangtze University
No. 1800, Lihudadao Road, Wuxi, Jiangsu 214122, China
{wxfangwei, sunjun_wx, xwb_sytu}@hotmail.com

Abstract. Design IIR digital filters with arbitrary specified frequency is a multi-parameter optimization problem. In this paper, we employ our proposed method, Quantum-behaved Particle Swarm Optimization (QPSO), to solve the IIR digital filters design problem. QPSO, which is inspired by the fundamental theory of Particle Swarm Optimization and quantum mechanics, is a global convergent stochastic searching technique. The merits of the proposed method such as global convergent, robustness and rapid convergence are demonstrated by the experiment results on the low-pass and band-pass IIR filters.

1 Introduction

Digital IIR filters are widely used in the fields of automatic control, telecommunications, speech processing and etc. There are two main methods for IIR digital filters design [1]. Recently, many researchers have devoted themselves to the optimization techniques and proposed them for IIR digital filters design [2], [3], [4], [5]. Quantum-behaved Particle Swarm Optimization (QPSO), which was proposed by Sun [6], [7], is a novel algorithm inspired by fundamental theory of particle swarm and quantum mechanics. The algorithm has strong search capability and is superior to the PSO [10] and Genetic Algorithm (GA).

The rest of the paper is organized as follows: In section 2, PSO is introduced and QPSO is detailed described and analyzed. Section 3 gives a brief introduction to the IIR digital filters and discusses how to apply QPSO in the design of IIR digital filters. Section 4 shows the experiment results and at last some conclusion remark is present.

2 Quantum-Behaved Particle Swarm Optimization

PSO, originally proposed by J.Kennedy and R.Eberhart [8], is a new global search technique. In PSO, each individual is treated as a volume-less particle in the D-dimensional space, with the position and velocity of *i*th particle represented as $X_i = (X_{i1}, X_{i2}, \cdots, X_{iD})$ and $V_i = (V_{i1}, V_{i2}, \cdots, V_{iD})$. The particles move according to the following equation:

$$V_{id} = w * V_{id} + c_1 * r_1 * (P_{id} - X_{id}) + c_2 * r_2 * (P_g - X_{id}), \quad X_{id} = X_{id} + V_{id} \tag{1}$$

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 637 – 640, 2006.

where c_1 and c_2 are acceleration coefficients, r_1 and r_2 are two random numbers. Vector $P_i = (P_{i1}, P_{i2}, \cdots, P_{iD})$ is the best previous position (the position giving the best fitness value) of particle *i* called *pbest*, and vector $P_g = (P_{g1}, P_{g2}, \cdots, P_{gD})$ is the position of the best particle among all the particles in the population and called *gbest*. Parameter *w* is the inertia weight introduced to accelerate the convergence speed of the PSO.

Quantum-behaved Particle Swarm Optimization, inspired by quantum mechanics, is a novel variant of original PSO with the following basic iteration equation,

$$x(t+1) = p \pm [L(t) * \ln(1/u)]/2, \quad \mathrm{u} = \mathrm{rand}(0,1), \tag{2}$$

where *u* is a random number distributed uniformly on (0,1), *p* is a stochastic point between the previous optimal position of the particle P_i and the global best position P_g, i.e. $p = (P_i \cdot \varphi_1 + P_g \cdot \varphi_2)/(\varphi_1 + \varphi_2)$, where φ_1 and φ_2 are random number in [0 1]. The value of *L* is calculated as $L(t) = 2 * \beta * | mbest - x(t) |$ [10], where

$$mbest = \frac{1}{M}\sum_{i=1}^{M} P_i = (\frac{1}{M}\sum_{i=1}^{M} P_{i1}, \frac{1}{M}\sum_{i=1}^{M} P_{i2}, \cdots, \frac{1}{M}\sum_{i=1}^{M} P_{id},) \cdot \tag{3}$$

Therefore, equation (2) can be written as [10]:

$$x(t+1) = p \pm \beta * | mbest - x(t) | * \ln(1/u) . \tag{4}$$

3 Application of QPSO to the Design of Digital IIR Filter

The frequency response of second order IIR digital filters is

$$H(e^{jw}) = A \cdot \prod_{k=1}^{L} \frac{1 + b_{1k} e^{-jw} + b_{2k} e^{-2jw}}{1 + a_{1k} e^{-jw} + a_{2k} e^{-2jw}} = A \cdot G(e^{jw}), \tag{5}$$

where

$$G(e^{jw}) = \prod_{k=1}^{L} \frac{1 + b_{1k} e^{-jw} + b_{2k} e^{-2jw}}{1 + a_{1k} e^{-jw} + a_{2k} e^{-2jw}} \tag{6}$$

Let us assume that the desired amplitude response of the filter on a set of discrete frequencies $\{ w_i, i = 1,2,\cdots,M \}$ is $H_d(e^{jw_i})$. Now the main design problem is to determine the coefficients in the equation (5) so as $H(e^{jw_i})$ approximates $H_d(e^{jw_i})$ as closely as possible. This approximation can be achieved by minimizing

$$E = E(\phi, A) = \sum_{i=1}^{M} [| A_0 \cdot G(e^{jw_i}, \phi) | - | H_d(e^{jw_i}) |]^2 , \tag{7}$$

where $\phi = [b_{11}, b_{21}, a_{11}, a_{21}, \cdots, b_{1L}, b_{2L}, a_{1L}, a_{2L}]^T$ and gain A_0 is given as follows[1]:

$$|A_0| = [\sum_{i=1}^{M} |G(e^{jw_i}, \phi)| \cdot |H_d(e^{jw_i})|] / [\sum_{i=1}^{M} |G(e^{jw_i}, \phi)|^2]. \quad (8)$$

In QPSO, the position of each particle represents the solution of the actual problem. Then the quality of the solution is dependent on the position of the particles. According to the characteristics of QPSO, the position of each particle is real-coded with each component representing a parameter of the designed filter. In order to ensure the stability of the filter, we set the same restriction to the parameters just like in [3], [4], [5].

4 Experiments and Results

We take the same experiments in [3], [4], [5]. Each algorithm runs 100 times independently. The three algorithms take the same population size and iteration which are 50 and 500; In PSO, maximum velocity is 4; In GA, tournament selection, uniform crossover and uniform mutation are used and Crossover Rate=0.8, Mutation Rate=0.05.

Experiments: design a low-pass and a band-pass IIR digital filter and the frequency response is as follows and the number of frequency sampling data is 46 and 50:

$$H_d(e^{jw}) = \begin{cases} 1 & 0 \le \omega \le 0.4\pi \\ 0.5 & \omega = 0.45\pi \\ 0 & 0.5\pi \le \omega \le \pi \end{cases}, \quad H_d(e^{jw}) = \begin{cases} 1 & 0 \le \omega \le 0.28\pi, 0.72\pi \le \omega \le \pi \\ 0.5 & \omega = 0.3\pi, \omega = 0.7\pi \\ 0 & 0.32\pi \le \omega \le 0.68\pi \end{cases}.$$

From Figure 1, it can be seen that the filters designed by QPSO has less ripple in the pass band and larger decline in the stop band. In Figure 2, it is apparent that QPSO can converge rapidly and get more good results than that of GA and PSO. Table 1 shows that QPSO can generate better mean results with small variance than any other two algorithms, which testify the robustness of QPSO.

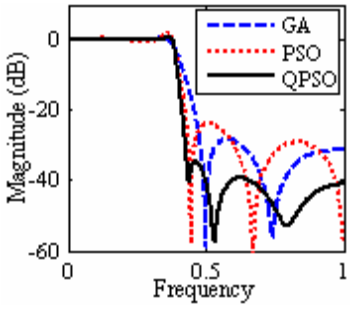


Fig. 1. Magnitude response of Low-pass IIR digital filter

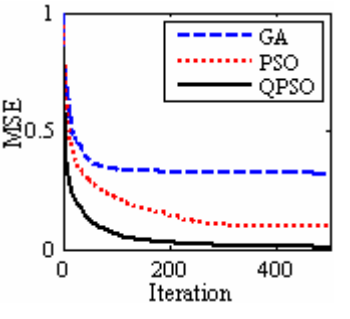


Fig. 2. Convergence behaviors in design Low-pass IIR digital filter

Table 1. Comparision of the experiment results

Filter Type	Algorithm	Mean Best	Variance	St.Dev.
Low-pass IIR digital filter	GA	0.3151	0.0204	0.1431
	PSO	0.0944	0.0264	0.1626
	QPSO	0.0079	8.6838E-5	0.0093
Band-pass IIR digital filter	GA	0.8894	0.0522	0.2285
	PSO	0.7608	0.2822	0.5313
	QPSO	0.3603	0.0118	0.1086

5 Conclusions

In this paper, with optimizing IIR digital filters being a multi-parameter problem, applying QPSO to the design of IIR digital filters has been studied. The proposed algorithm is global convergent, robustness and rapid convergent. From the experiment results it was concluded that QPSO could be efficiently used for IIR digital filter design and superior than GA and PSO.

References

1. Chen Peiqing.: Digital Signal Processing 2nd ed. Beijing: Tsinghua University Press. (2001)
2. Hu Guangshu.: Digital Signal Processing: Theory, Algorithm and Realization, 2nd ed. Beijing Tsinghua University Press. (2003)
3. Li Jianhua, Yin Fuliang.: Genetic Optimization Algorithm for Designing IIR Digital Filters. Journal of China Institute of Communications, 17 (1996) 1-7
4. Hou Zhirong, Lu Zhensu.: Particle Swarm Optimization for IIR Digital Filters Design. Journal of Circus and Systems, 8 (2003) 16-20
5. Zhang Gexiang, Jin Weidong, Hu Laizhao.: An effective Optimization Method for Designing IIR Digital Filters. Signal Processing, 20 (2004) 152-156
6. J. Sun et al.: Particle Swarm Optimization with Particles Having Quantum Behavior. Proceedings of Congress on Evolutionary Computation. (2004) 325-331.
7. J. Sun et al.: A Global Search Strategy of Quantum-behaved Particle Swarm Optimization. Proceedings of IEEE Conference on Cybernetics and Intelligent Systems. (2004) 111-116
8. J. Kennedy, R. Eberhart.: Particle Swarm Optimization. Proceedings of IEEE Conference On Neural Network. (1995) 1942-1948
9. Shi Y, R, Eberhart.: A Modified Swarm Optimizer. Proceedings of IEEE Conference on Evolutionary Computation. (1998) 69-73
10. J.Sun, Wenbo Xu and Jing Liu.: Parameter selection of Quantum-behaved Particle Swarm Optimization, Proceedings of International Conference of Natural Computing. (2005) 543-552

Optimization of Finite Word Length Coefficient IIR Digital Filters Through Genetic Algorithms – A Comparative Study

Gurvinder S. Baicher

Department of Engineering, Newport Business School, University of Wales Newport, Allt-yr-yn Campus, P.O.Box 180, Newport NP20 5XR, South Wales, U.K.
g.singh-baicher@newport.ac.uk

Abstract. This paper considers the specific issues relating to finite word length (FWL) coefficient constraints for the case of infinite impulse response (IIR) digital filters. Due to the feedback nature of recursive filters, stability issues are an important factor in their design and are discussed in some detail. Some previously reported work on the optimization of FWL coefficients for IIR filters is also discussed. Extensive range of filter types and structures of IIR filters and their optimization using genetic algorithms is investigated and reported. Finally, comparative tests were conducted using the simple hill climber optimization techniques for a selection of filters.

1 Introduction

In many examples of real-time; low power and high throughput applications, the 'infinite precision' coefficients of digital filters derived from their design have to be replaced by finite word length equivalents. Simplistically, this can be achieved by rounding the coefficients to their specified word length. However, such a truncation can result into an unpredictable frequency response that may become unacceptable. For this reason, optimization techniques are commonly used to derive a variant of the set of finite word length coefficients for near optimal design.

The finite word length coefficient optimization of finite impulse response (FIR) digital filters has been previously reported [1]. In this paper we consider the case of infinite impulse response (IIR) digital filters. The exact design and analysis of IIR filters is normally based in terms of linear systems. However, when finite word length effects of quantization error and overflow are considered then the system becomes non-linear and it is this that causes difficulties in the analysis of fixed point filter implementation. For recursive filter structures the problems of the effects of finite word length become more severe when compared to the non-recursive filters. In an extreme situation and especially for narrow band filters where the poles of the filter are fairly close to the unit circle, the finite word length coefficients of IIR filter may generate positive feedback and thus become unstable. The finite word length realization of recursive IIR filters due to fixed-point hardware suffer from the same error effects as for non-recursive FIR filters such as; input signal quantization due to analogue to digital conversion, coefficient quantization, overflow errors and product

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 641 – 650, 2006.

round-off errors. In addition to these errors, two additional problems are caused by the feedback nature of the recursive filter. These are; firstly, small-scale limit cycles that are oscillations caused by quantization non-linearity in the feedback loop and secondly, large scale limit cycles caused by fixed-point arithmetic non-linear overflow in the feedback loop. Of these two problems, the first error is usually low and is easily tolerated but the second error can lead to large-amplitude sustained oscillations over the complete dynamic range of the recursive filters and so must be prevented.

The commonly used structures for IIR digital filters are; direct, parallel and cascade forms. It is well established that a cascade of 2^{nd} order sections is much less sensitive to coefficient quantization effects and its impact on stability of the filter, especially when compared to direct form implementation [2], [3]. Proper ordering and matching of poles and zeros in each section of the cascade and their scaling further ensures reduced coefficient sensitivity [4], [5]. Other structures that can be used for IIR filter realization are: state-variable [6], lattice [7] and the wave digital filter [8]. Generally, as the filter structure becomes more complex then a larger part of the process of filtering is carried by the structure itself placing less load on the coefficients. Thus, more complex structures such as lattice and wave digital filters, are capable of operating to a required response, with greatly reduced coefficient word lengths. However, in terms of coefficient sensitivity, the 2^{nd} order cascade has been shown to perform well even when compared to parallel and wave digital filter structure for a number of designs [9]. Hence, in this study, 2^{nd} order cascades are used due to their simplicity of filtering algorithms and ease of performing stability tests.

2 Optimization and Stability Issues in IIR Filter Design

In a recursive digital filter the output is a linear combination of past and present inputs and past outputs. The difference equation representation is of the form

$$y(n)= \sum_{m=0}^{M} b_m x(n-m) - \sum_{m=1}^{N} a_m y(n-m) \tag{1}$$

and the transfer function is

$$H(z)= \frac{\sum_{m=0}^{M} b_m z^{-m}}{1+ \sum_{m=1}^{N} a_m z^{-m}} \tag{2}$$

The coefficients b_m and a_m of the filter transfer function H(z) of eqn. (2) are obtained in a high precision form through the initial design stage of the IIR filter such that the stipulated filter specifications are satisfied. For real-time realization of such filters using fixed-point devices, then errors arise in a number of ways that could degrade the performance of the filter and in extreme situations make an otherwise stable filter to become unstable.

Appropriate scaling of the inputs to the adders such that the outputs are kept low can prevent the large-scale limit cycles. This can lead to a reduction of the signal-to-noise ratio (SNR) of the system. It is thus important to select an appropriate scaling

factor for a given structure to prevent overflow while at the same time preserving the best possible SNR. There are three commonly used methods for determining a suitable scaling factor and are referred to as the L_1, L_2 and L_∞ norms [10].

The issue concerning coefficient sensitivity and filter stability has been analyzed by considering the extent to which the pole positions change as a result of changes in the coefficient a_m [3], [11]. Some of the important deductions of the analysis are

- The filter is most sensitive to the last coefficient a_N.
- Moving the poles closer to the unit circle increases the sensitivity of the pole location to the variation of a coefficient.
- Coefficient sensitivity increases when the poles are close together.

For high order filters, the poles are normally clustered together in the pass band, so in order to reduce sensitivity a cascade structure of several lower order sections is recommended. This leads to the preference for using a cascade of 2nd order section. By factoring the rational transfer function of eqn. (2), the cascade 2nd order transfer function can be written as

$$H(z) = \prod_{k=1}^{L} \frac{b_{0k} + b_{1k} z^{-1} + b_{2k} z^{-2}}{a_{0k} + a_{1k} z^{-1} + a_{2k} z^{-2}} \tag{3}$$

where L is the number of cascade sections.

There are many different cascade structures possible depending on the ordering of $H_k(z)$ blocks and also on different pairings of the poles and zeros. This flexibility of structural leniency allows for a choice of filter realizations that may be used to reduce coefficient quantization dependency. For reduced coefficient sensitivity, a specific 2nd order cascade design procedure is described in [5]. Some important considerations in the ordering of the cascade sections are

- Match the poles closest to the z-plane unit circle with the zeros closest to those poles.
- Match the remaining poles to zeros similarly, moving towards the origin of the unit circle.
- The section with poles closest to the unit circle is ordered as the last section of the cascade preceded by other pole/zero pairing sections according to the distance of the poles from the unit circle.

The 3rd condition given above is based on the assumption that the following scaling of the sections in the cascade is applied i.e.

$$\text{Max} \left| \prod_{i=1}^{N} H_i(z) \right| < 1 \qquad N=1,\ldots.., L-1 \tag{4}$$

The scaling property of equ. (4) generates maximum peaking of the magnitude response of the section with poles closest to the unit circle. The ordering rule, therefore, is to start with sections that are least peaked and move towards the most peaked.

The Matlab Signal Processing toolbox provides a function tf2sos (transfer function to second order section convert) that can be used to generate second order sections that are ordered according to the above rules based on appropriate optimum pole-zero

pairing. This function generates a matrix sos containing the coefficients of each second order section of the equivalent transfer function H(z).

3 GA Optimization of IIR Digital Filters

Genetic algorithms are a useful tool in the optimization of finite word-length constrained coefficients for realization of IIR digital filters. Some examples of GA optimization applied to the IIR filter design problems have been reported in literature. Wilson and Macleod [12] consider a cascade form design of IIR filters. A simple GA is applied in order to find a compromise solution to the frequency response error and adder cost. In addition, stability and minimum phase is guaranteed by analyzing the genes and identifying positions of poles and zeros. If the root is outside the unit circle then this is moved by multiplication with an all pass filter. Quantizing the coefficients then follows this procedure. This step is a restriction to direct optimization of the realizable IIR filter. Harris and Ifeachor [13] have considered an automated design procedure for IIR filters. This work considers a hybrid GA approach that optimizes second order cascade sections of IIR filters in terms of pole-zero positions on the z-plane. The GA culminates when an appropriate filter is located within specified bounds of maximum pass-band ripple and minimum stop band attenuation. The stability criterion of the stability triangle is used to return unstable solutions with low fitness function. A multiple objective fitness function includes a weighted component of round-off noise due to the ordering of the second order sections and the frequency response of the filter. This combination allows a compromise solution to be found based on the two variables that can be controlled by the design specifications. Arslan and Horrocks [14] have reported other work in this field. Again, second order sections are considered here, using real-valued coefficients that are arranged in a concatenated form represented by a string of cascade stages. The frequency response template is specified within minimum and maximum limits and the overall fitness is evaluated as a function of the deviation from the exact design frequency response of the filter. Stability checks are also conducted based on the stability triangle.

The main advantage of the work covered here and reported in this study is that the complete process for design and optimization is Matlab based and can be executed sequentially by selecting the appropriate choice of the filter design and the optimization parameters. This option offers flexibility of design using the standard Matlab Signal Processing toolbox functions that include issues about scaling, pole-zero pairing and ordering of the second-order sections. The GA used in this work is a Matlab Toolbox developed by Chipperfield et al [15]. This toolbox was originally developed for Control Systems applications and has been adapted for IIR digital filter optimization. The main GA functions such as ranking, crossover and reinsert were used without any change and new m-file function for calculating the fitness function was written.

Such a Matlab based integral approach to the initial design of the IIR filter and subsequent GA optimization makes this procedure flexible to obtaining quantifiable metrics for a number of design specifications. For example, the Matlab function tf2sos converts the high order rational transfer function H(z) into its second order sections in a cascade. The pole-zero ordering in the default option ‘UP’ is such that the first row of the matrix sos of equ. (5) will contain poles closest to the origin and

the last row will contain poles closest to the unit circle. This option allows for minimum coefficient sensitivity due to quantization. The 'SCALE' option of the tf2sos function specifies the desired scaling of the gain and the numerator coefficients of all the second order sections in the cascade. The 'SCALE' options available are the L_2-norm, L_∞-norm and no scaling. A combination of the default 'UP' ordering of pole-zero pairs and L_∞-norm scaling minimizes the probability of overflow error in the realization of the IIR filter.

The complete process of optimization is achieved in a single stage implementation using a single frequency response template of the exact magnitude response. The magnitude error function is calculated using the following

$$\text{mag\_error} = \sum_{i=1}^{L} \left| H_{ei} - H_i \right| \tag{6}$$

where, H_{ei} = Magnitude response of the exact (i.e. high precision coefficient) filter at frequency 'i' and H_i = Magnitude response of the test filter at frequency 'i'

The phase error function is similarly calculated using the phase response of the filter. The object function value that must be minimized is evaluated using

$$\text{Obj\_Val} = \text{mag\_error} + W \;\; \text{phase\_error} \tag{7}$$

where W = weighting fraction.

The percentage error between the exact and the best optimized magnitude response is calculated using the following

$$\%\ \text{mag\_error} = \frac{\sum_{i=1}^{L} \left| H_{ei} - H_i \right|}{\sum_{i=1}^{L} \left| H_{ei} \right|} \; 100 \tag{8}$$

4 Test IIR Filters and GA Optimization Results

A number of example IIR filters were used to test the robustness of the GA optimization code. The filters are tested for the direct form format and the second-order cascade form for a number of different types, order and number of bits. Table 1 shows the different types of filters used i.e. A-type is a low pass filter with cut-off frequency Wn = 0.5π, B-type is a band pass filter with cut-off frequencies Wn = [0.3π 0.7π] and C-type is high pass filter with cut-off frequency Wn = 0.6π. Each of these filters was tested for 4^{th}, 6^{th} and 8^{th} order using 5, 8 and 12 bits in each case. Further distinction of filters used were to test for minimal phase and non-minimal phase in the case of direct form filters and for the second order cascade form the case of 'Infinity norm' and 'No norm' options were both used. The GA specifications used are

Population size (Nind) =120
Maximum number of generations =20
Weighting fraction (W) = 0.001
Number of frequency points (L) = 500
Generation gap (GGAP) = 0.8
Insertion rate (INSR) = 1

Table 1. Set of IIR Filter Specifications Note that the 'ellip' function of Matlab specifies cut-off frequencies Wn between 0 and 1 where 1 is the Nyquist frequency

Filter	Wn
A: range (low pass) Pass band ripple: Stop band attenuation:	0.5 1 dB 40 dB
B: range (band pass) Pass band ripple: Stop band attenuation:	[0.3 0.7] 1 dB 40 dB
C: range (high pass) Pass band ripple: Stop band attenuation:	0.6 1 dB 40 dB

The results for direct form and 2nd order cascade form structure GA optimization of IIR filters for various types, order and number of bits used are comprehensively covered in [16]. Some examples of the GA optimized response for a selection of filters are shown in Figs. 1-6.

5 Simple Hill Climber Techniques and Exhaustive Search

To test the robustness and accuracy of the GA optimized results, the methods of simple hill climber algorithms such as the steepest ascent (SAHC) and the nearest ascent (NAHC) were applied to a selection of filters shown in Table 1 for the second-order cascade form structure of IIR filters. A flow chart for the hill climber algorithm used in this work is shown in [1]. Random sampling tests for the search space as used for the GA optimization were also conducted. Furthermore, for a selection of low order filters, an exhaustive search was conducted over a matching search space.

An important observation for the application of hill climber algorithms is that the search space for optimization can extend beyond the range of +1 or -1 of the rounded values for each coefficient. This outcome is implicit in the evolutionary nature of the algorithms since mutation of the coefficient value occurs for each iteration. In this

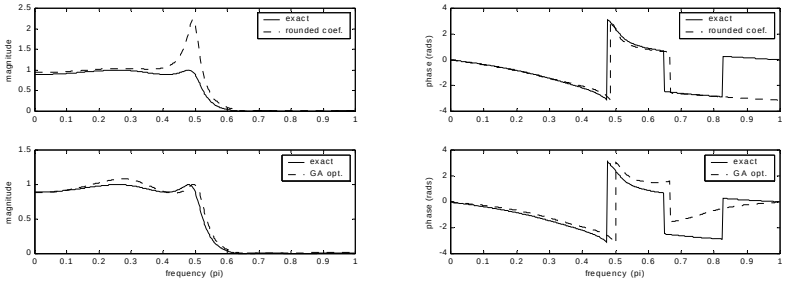


Fig. 1. Magnitude and phase response of a 4th order direct form low-pass filter using 5 bit coefficients and non-minimal phase

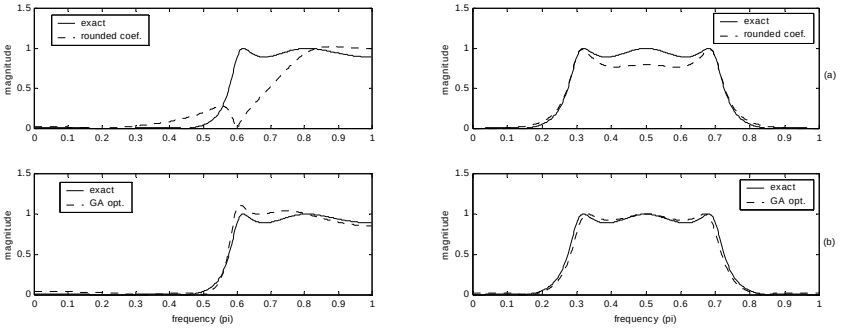


Fig. 2. Magnitude response of a 4th order direct form high-pass filter **(a)** and a 6th order direct form band pass filter **(b)** using 5 bit coefficients

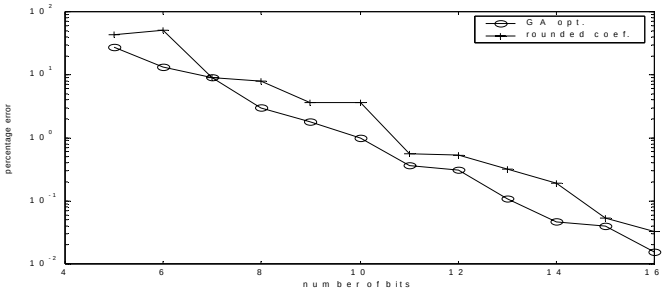


Fig. 3. Magnitude response percentage error against number of bits of a 6^{th} order direct form low-pass filter and non-minimal phase

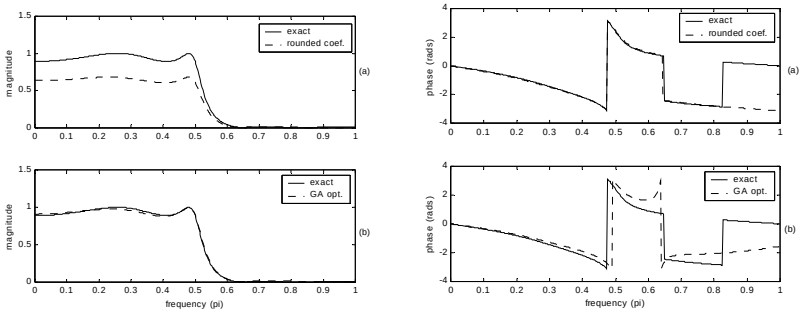


Fig. 4. Magnitude and phase response of a second order cascade form 4^{th} order low-pass filter using 5 bit coefficients and Infinity norm

respect, there is a subtle difference when compared with the GA optimization because the search space for GA is confined to +1 and -1 of the rounded coefficient values for the results obtained in this study.

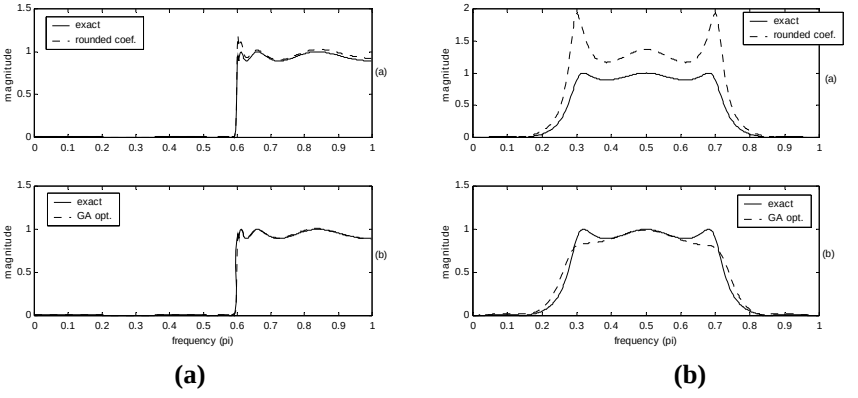


Fig. 5. (a) Magnitude response of a second order cascade form 8th order high-pass filter using 8 bit coefficients. **(b)** Magnitude response of a second order cascade form 6th order band-pass filter using 5 bit coefficients.

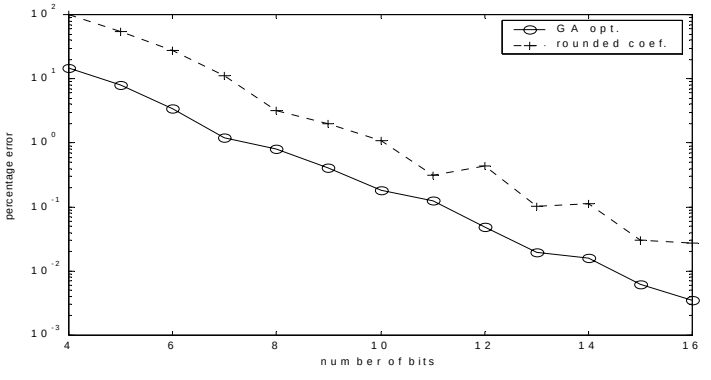


Fig. 6. Magnitude response percentage error against number of bits of a second order cascade form 8th order high-pass filter with 'None' norm

The hill climbers are thus subjected to a wider search space that may or may not be advantageous to the optimization process. The results of SAHC, NAHC, the random sampling and exhaustive search for a selection of the IIR filters are shown in Table 2. The results shown with an asterisk (*) are the ones for which the search space has deviated greater than +1 or -1 of the rounded coefficient values. Note also that the exhaustive search was confined to deviation of +1, 0 or –1 of the rounded coefficients.

Table 2. GA and hill climber optimization results of summed magnitude error of equ. (6) over 500 frequency points for second order cascade form IIR filters.

Filter	Exh. Sch	Random	SAHC	NAHC	GA
1) NN/SOS/LP4/5	6.1447	6.1447	9.3285	30.3495	6.1447
2) NN/SOS/LP4/8	0.9172	0.9253	0.8200*	2.2104	0.9172
3) IN/SOS/LP4/8	0.7966	0.7966	0.8812*	1.7215*	1.0973
4) IN/SOS/HP4/5	9.2286	11.5684	11.0276	21.6502	10.3195
5) NN/SOS/HP4/8	1.1778	1.4768	1.0779*	3.9608*	1.4458
6) NN/SOS/BP4/5	14.9344	22.0861	14.9344	52.2511	14.9344
7) IN/SOS/BP4/8	3.1483	3.1483	3.2204	5.0610	3.1483

Exh. Sch = Exhaustive search; Randon=random sample; SAHC=steepest ascent hill climber NAHC=nearest ascent hill climber

* indicates search space exceeded +1 and/or –1 variation of rounded coefficient values

NN= 'none' norm, IN= 'infinity' norm, LP4/5= 'low' pass 4^{th} order using 5 bits

6 Discussion of Results and Conclusions

As no quantifiable metrics for the example form and structure of IIR filters and their FWL optimized coefficients are available in literature, an extensive range of such filters have been arbitrarily selected and used in this study. These cover both the direct form and the second order cascade form structures. An investigation of GA optimization for each type of the test filters was conducted and a selection of optimized frequency responses is reported. It must be recognized that for stability of IIR filters, all poles of the transfer function must lie inside the unit circle of the 'z' plane. This condition is embedded in the GA optimization code thus leading to results that assure stable filter design. The GA optimization was conducted for the case of minimum phase (MP) and non-minimum phase (NMP) conditions of IIR direct form filters in case of the example filter specifications. The results of GA optimization show a significant improvement over the simply rounded coefficient values.

Further tests were conducted on a selection of IIR filters of the second order cascade form structures using the simple hill climber techniques, random sampling and exhaustive search. The results of these tests are shown in Table 2. Once again, the GA optimized results are seen to be consistently good. However, for some filters such as the Filter2: NN/SOS/LP4/8, Filter3: IN/SOS/LP4/8 and the Filter 5: NN/SOS/HP4/8, the steepest ascent hill climber method has generated superior results. This is significant since the search space for these algorithms can intrinsically extend beyond the +1 and/or –1 of the rounded coefficient values. The GA search space, however, is restricted to +1 or –1 of the rounded coefficients. This offers credibility to the simple hill climber technique and complements the GA optimization to search for superior solutions for the application considered here.

References

1. Baicher, G.S., Taylor, M. and Rowlands, H.: A comparative study of finite word length coefficient optimization of FIR digital filters, Proceedings of the International Conference on Natural Computation (ICNC–05), Changsha, Hunan, China. Lecture Notes in Computer Science, Springer. Vol. 3612, pp 877-882, 27-29 August (2005).

2. Kaiser, J.F.: Some practical considerations in the realization of linear digital filters. Proceedings of the 3rd Annual Allerton conference on Circuit System Theory. pp 621-633 (1965).
3. Parks, T.W. and Burrus, C.S.: Digital Filter Design. New York: John Wiley, (1987).
4. Oppenheim, A.V and Shaffer, R.W.: Discrete-time Signal Processing. Englewood Cliffs, N.J: Prentice-Hall (1989).
5. Jackson, L.B.: Digital filters and signal processing. Boston: Kluwer (1989).
6. Roberts, R.A and Mullis, C.T.: Digital Signal Processing. Reading, MA: Addison-Weseley (1987).
7. Gray, A.H. and Markel, J.D.: Digital Lattice and Ladder filter synthesis. IEEE Transactions on audio and electroacoustics. vol. **21**, pp 491-500, (1973).
8. Fettweis, A.: Wave digital lattice filters. International Journal on Circuit theory application. vol. **2,** pp 203-211, (1974).
9. Dempster, A.G. and Macleod, M.D.: Multiplier blocks and the complexity of IIR structures. IEE Electronics letters. vol. **30**, pp. 1841-1842 (1994)..
10. Ifeachor, E.C. and Jervis, B.W.: Digital Signal Processing: A practical approach. Wokingham, England: Addison-Wesley publishing (1993).
11. Mitra, S.K.: Digital Signal Processing – a computer based approach. New York: McGraw-Hill (1998).
12. Wilson, P.B. and Macleod, M.D.: Low implementation cost IIR digital filter design using genetic algorithms. Proceedings of the workshop on "Natural algorithms in signal processing". vol.1, pp.4/1-4/8, Chelmsford, U.K., Nov. (1993).
13. Harris, S.P. and Ifeachor, E.C.: Automatic design of frequency sampling filters by hybrid genetic algorithm techniques. IEEE Transactions on Signal Processing. vol. **46**(12), pp. 3304-3314 (1998).
14. Arslan, T. and Horrocks, D.H.: A genetic algorithm for the design of finite word length arbitrary response cascaded IIR digital filters. Proceedings of IEE International conference on Genetic Algorithms in Engineering Systems: Innovations and Applications (GALESIA, 1995), pp.276-281, (1995).
15. Chipperfield, A., Fleming, P., Pohlheim, H. and Fonseca, C.: Genetic Algorithm Toolbox for use with MATLAB. Dept. of Automatic Control and Sys. Eng., University of Sheffield, ver. 1.2, (1993).
16. Baicher, G.S.: Towards Optimisation of Digital Filters and Multirate Filter Banks through Genetic Algorithms, PhD Thesis, University of Wales, Newport, South Wales, U.K. (2003).

A Computer Aided Inbetweening Algorithm for Color Fractal Graphics

Yunping Zheng[1], Chuanbo Chen[1], and Mudar Sarem[2]

[1] College of Computer Science and Technology, HUST, 430074, Wuhan, Hubei, China
{zypdragon, chuanboc}@163.com
[2] School of Software Engineering, HUST, 430074, Wuhan, Hubei, China
mudar66@hotmail.com

Abstract. The key frame technology is one of the most important technologies in the computer animation. All the researchers in the key frame algorithm are mainly focusing on the Euclidean object, without considering the fractal object. Fractal geometry can describe many irregularly shaped objects or spatially non-uniform phenomena in nature that cannot be accommodated by Euclidean geometry. In addition, the fractal objects in nature are almost colorful. Therefore, a computer aided inbetweening algorithm for color fractal graphics is presented in this paper. Our study shows that the fractal idea can be effectively applied in the key frame animation. The main feature of our algorithm is that it can deal with a fractal object while the conventional algorithm cannot. The theoretical and experimental results presented in this paper show that our algorithm has many practical values that can improve the efficiency of animation production and simultaneously greatly reduce the cost.

1 Introduction

Inbetweening is an important process in traditional 2D animation, where animators draw a sequence of key frames firstly, and then assistant animators draw in-between frames correspondingly. During this process animators use their experience and semantic knowledge of nature of objects and their motion. Usually a huge number of in-betweens have to be drawn manually during the generation for the animation. However, an automatic inbetweening system could cut down significant amount of time and labor. Automatic inbetweening is one of the main focuses in computer assisted animation as inbetweening process is traditionally time-consuming and labor-intensive in the production of cartoon animation. One strength of key frame animation techniques is the analogy to conventional hand animation techniques, simplifying the transition when a classically trained animator adapts to using computers. For the time being, all the researchers in the key frame algorithm are mainly focusing on the Euclidean object, without considering the fractal object. In fact, Mandelbrot [1] has shown that mountains, clouds, aggregates, galaxy clusters, and other natural phenomena are similarly fractal in nature, and fractal geometry's application in the sciences has become a rapidly expanding field. Fractals are used especially in computer modeling of irregular patterns and structures in nature. Fractals are not relegated exclusively to the realm of mathematics. If the definition is broadened a bit, such objects can be found virtually everywhere in the natural world.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 651 – 659, 2006.

2 Related Work

The key frame technology is one of the most important technologies in the computer animation. Xiaogang et al. [2] only studied the Euclid objects among the computer aided inbetweening algorithms. Seah and Feng [3] proposed a method based upon using both distance vector estimation and segmentation of images in regions. Fekete et al. [4] presented an approach to the whole vectorized drawing process by providing an interface for a paperless system. Kort [5] offered an interactive model for auto inbetweening with a possible solution for "undetectable" parts. Rademacher [6] proposed a solution for viewer-dependent geometry for characters already modeled in three dimensions. Except for the 2D and 3D animation techniques, there is also a 2.5 D technique [7]. In [8], we put forward a key frame algorithm by meshing a three-dimension object.

However, all of the key frame approaches above are unable to deal with the fractal objects which can only be described by fractal geometry, because they strongly depend on the underlying geometric model which can only be described by conventional Euclidean geometry. The conceptual roots of fractals can be traced to the attempts to measure the size of objects for which traditional definitions based on Euclidean geometry or calculus failed. Thus, the objects of the key frame animation should not be restricted to Euclidean objects. In addition, all of the fractal objects in nature are almost colorful and the beauty of fractals has made them a key element in computer graphics. Therefore, in this paper, a new computer aided inbetweening algorithm for color fractal graphics is presented.

The remainder of this paper is organized as follows: In section 3, some preliminaries used in our algorithm are reintroduced. We present a novel fractal-based animation for color fractal graphics in section 4. Then, a computer aided inbetweening algorithm for color fractal graphics is put forward in section 5. Next, we test our algorithm with the Matlab 6.5 software and analyze experimental results in section 6. Finally, in section 7, we conclude our work and discuss the future work.

3 Preliminaries

In this section, the first three theorems which have been introduced in [9] are represented, since this study is highly depends on them in designing our new algorithm. The fourth theorem, which we put forward in this paper, can guarantee that each inbetween is still a fractal during the inbeweening.

Theorem 1. Suppose an affine transformation $W : R^2 \to R^2$ is a similitude as $W(x) = Ax + t$ .If and only if $\rho(A) < 1$, the transformation W is contractive, where $\rho(A)$ denotes the spectrum radius of the matrix A.

Theorem 2. Suppose (X, d) is the metric space and $\{X; (W_0), W_1, W_2, ..., W_N\}$ is an IFS, the contractive factor $s = \max\{s_n : n = 0,1,...,N\}$, if each W_n continuously depends on the parameter $p \in P$, then the attractor $A(p) \in \Phi(X)$ continuously

depends on $p \in P$ according to the Hausdorf measure $h(d)$, where P is the metric space and $\Phi(X)$ is the set of all non-empty, compact subsets of X .

Theorem 2 shows that the attractor can be continuously controlled by adjusting some parameters of the transformation.

Theorem 3. Suppose $\{X; \omega_i, i = 1,2,...,n\}$ is an IFS and its attractor is A .If $\omega_k = \omega_l . \omega_m$ $(k,l,m \in \{1,2,...,n\})$, then the attractor of $\{X; \omega, i = 1,2,...,n, i \neq k\}$ is still A.

Theorem 3 can increase or decrease the number of the contractive mapping sets, but the attractor remains invariant.

Theorem 4. Suppose the corresponding affine transformation matrix of the contractive mapping sets $W_i, i = 1,2,...,m$ and $\psi_j, j = 1,2,...,n$ is A and B respectively, then the united matrix $M = (1 - f(t)) \times A + f(t) \times B$ is also a contractive matrix, where $f(t) \in [0,1]$ and $t \in [0,1]$.

Proof

(1) If $f(t) = 0$ or $f(t) = 1$, then M equals A or B, obviously M is a contractive matrix.

(2) If $f(t) \in (0,1)$, then $(1 - f(t)) \in (0,1)$, since both A and B are the contractive matrix, according to the theorem 1, we know $0 < \rho(A) < 1$ and $0 < \rho(B) < 1$, therefore $\lim_{n \to \infty} A^n = 0$ and $\lim_{n \to \infty} B^n = 0$ .Then according to the binomial expand theorem, we know $\lim_{n \to \infty} M^n = \lim_{n \to \infty} ((1 - f(t) \times A) + f(t) \times B)^n$,

$\lim_{n \to \infty} M^n = \lim_{n \to \infty} \sum_{r=1}^{n} C_n^r (1 - f(t) \times A)^{n-r} (f(t) \times B)^r = 0$. So $0 < \rho(M) < 1$, i.e. M is also a contractive matrix.

Theorem 4 guarantees that each in-between is still a fractal.

4 Fractal-Based Animation for Color Fractal Graphics

The essence of the inbetweening problem is determining the correspondences between the key frames. In a simplified view, inbetweening is the process of generating all the frames of a motion sequence given its start and end key frames.

The basic idea of IFS is that a fractal is a shape made of parts similar to the whole in some way [1]. It is the self-similarity that makes it possible to generate a graphic by the iteration. From the start of one point or a simple geometric graphic, a complex graphic can be produced by the iteration according to some regulars. At last, the acquired target point sets have nothing with the initial given point and it only depends on the regulars of the iteration.

In the traditional key frame algorithms, the research object is the Euclidean object. The basic ideas in these traditional algorithms are consistent. Initially, there are three steps in the key frame process: the matching mapping, the interpolation of the vertexes, and the rendering method. However, these traditional algorithms don't fit for the fractal object, because in the fractal geometry the point, the line, and the surface of an object are treated as a whole, not as a part.

As for the fractal-based key frame algorithm for color objects, there are three main problems which need to be solved: First, the corresponding principles of the contractive mappings, i.e., how to correspond each of the contractive maps between the start and the end key frame. Second, the corresponding principles of the probabilities, i.e., how to correspond each of the probabilities between the start and the end key frame. Third, the corresponding principles of the colors.

With regard to these three problems, a relatively good solution is presented in the following section 5.

5 The Description of Our Algorithm for Color Fractal Graphics

As a start point to describe our algorithm, we first introduce the interpolation function and the corresponding principle, which are the core part of our algorithm.

5.1 Interpolation Function

The interpolation function is an important function in producing the in-betweens. Therefore, it is essential to select an appropriate interpolation function to obtain high quality animation. In practice, we choose interpolation function f(t) which makes the in-between function G(t) meet $G(t)=(1-f(t))*G_b+f(t)*G_e=G_b+(G_e-G_b)*f(t)$ where the function f (t) meets the these properties: $f(0)=0$, $f(1)=1$, $0 \leqslant f(t) \leqslant 1$, $0 \leqslant t \leqslant 1$.

Among this kind of interpolation functions f (t), t is a pace factor controlling in-between interpolation. As long as the selected function meets the above conditions, the generated in-betweens will not have dramatic distortion.

5.2 Principium of Selecting the Corresponding Principle

Three corresponding principles about contractive affine transformations, probabilities, and colors of the start key frame and the end key frame are used in our algorithm. Firstly, if the number of the contractive mapping between the start and the target object is not equal, then by using theorem 3 we can increase or decrease the number until they are equal. Secondly, we normalize the probability for the start or the end key frame. Thirdly, according to the value of probabilities which are arranged in the degressive order, we can rearrange the subscript of all the contractive mappings and its corresponding color mappings in the ascending order for G_b and G_e respectively. Herein we can guarantee that the subscription of the contractive mapping and its corresponding color mapping is minimal whereas its probability used is maximal, and vice versa. Finally, according to the order of the subscription of the contractive mapping or its corresponding color mapping (degressive or ascending), the global mapping between G_b and G_e can correspond one by one.

On the contrary, if we don't adopt this kind of corresponding principles, we might as well suppose the number of the contractive mapping between the start and the target object is equal. Firstly, according to the value of probabilities which are arranged in the degressive order, we can rearrange the subscript of all the contractive mappings and its corresponding color mappings in the degressive order for G_b but in the ascending order for G_e respectively. Secondly, according to the order of the subscription of the contractive mapping or its corresponding color mapping (degressive or ascending), the whole mapping between G_b and G_e can correspond one by one. Finally, compute $P_k = p_k(1\text{-}f(t))+q_k f(t)$ and $M_k = (1- f(t))* B_k + f(t)* E_k$. Later on, it can lead to such fact that the contractive mapping whose probability is maximal (or minimal) before inbetweening is not maximum (or minimum) during inbetweening. As a result, the speed of each generating branch of the in-between is not greatly uneven, because the point generated by a mapping with a bigger probability should have been denser, but now it is sparser because of using this kind of corresponding principles ,and vice versa. Therefore, according to such principles, the density contrast of the generated points on the fractal graphic must be very obvious, because the points in large area should have been produced, but now they have no opportunity to be produced, and vice versa.

In short, for contractive mapping whose total number is m, there are m! arrangement methods. The degressive order is used for the arrangement of probabilities in our algorithm. The other m!-1 arrangement methods may also be used, but just as what we have analyzed above, other arrangement methods can not guarantee the generated points on a fractal graphic are even. For this reason, the corresponding principle put forward in the paper is obviously better than the rest.

5.3 The Steps of Our New Algorithm for Color Fractal Graphics

Without loss of generality, we will assume that inbetweening takes place over the interval of time [0, 1]. In practice, N denotes the number of control frames that are used to generate a motion sequence.

In fractal geometry, we might as well suppose the start and end key frame are G_b and G_e respectively and their contractive affine transformations are $W_i, i = 1,2,...,m(m \geq 2)$ and $\psi_j, j = 1,2,...,n(n \geq 2)$, the probability used by each transformation is p_i and q_j respectively, and color used by each transformation is B_i and E_j respectively. In this paper, the color is denoted by RGB system and let $B_i = \left[\overline{r_i}, \overline{g_i}, \overline{b_i}\right]$ and $E_j = \left[\overline{r_j}, \overline{g_j}, \overline{b_j}\right]$, where $\overline{r_i}, \overline{g_i}, \overline{b_i}$ and $\overline{r_j}, \overline{g_j}, \overline{b_j}$ are defined in the interval of [0,255].

Suppose the contractive mapping sets for G_b and G_e are:

$$W_i\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \overline{a_i} & \overline{b_i} \\ \overline{c_i} & \overline{d_i} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \overline{e_i} \\ \overline{f_i} \end{bmatrix} \text{ and } \psi_j\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a_j & b_j \\ c_j & d_j \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e_j \\ f_j \end{bmatrix} \text{ respectively.}$$

By the formula M= (1-f (t)) G_b+ f (t) G_e, where f (t) is the interpolation function, we can get the united contractive mapping sets which are used to generate the inbetweens:

$$\eta\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}\bar{a}_i(1-f(t))+a_i f(t) & \bar{b}_i(1-f(t))+b_i f(t)\\ \bar{c}_i(1-f(t))+c_i f(t) & \bar{d}_i(1-f(t))+d_i f(t)\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}+\begin{bmatrix}\bar{e}_i(1-f(t))+e_i f(t)\\ \bar{f}_i(1-f(t))+f_i f(t)\end{bmatrix}$$

The start key frame and the end key frame are given beforehand in our algorithm; we only use their IFS codes to generate the in-bewteens. The following steps describe our computer aided inbetweening algorithm for color fractal graphics.

Step 1. While m≠n, we might as well suppose m<n. According to theorem 3, then we can select a ω_i whose probability is maximum among p_i , let $\omega_{m+1}=\omega\times\omega$, $p_{m+1}=p_i\times p_i$, $B_{m+1}=\lambda * B_i$, then add them to $\{W_i, p_i, B_i; i=1,2,...,m(m\geq 2)\}$. Then we can get a new contractive mapping set $\{W_i, p_i, B_i; i=1,2,...,m+1(m\geq 2)\}$, where λ is a scale coefficient for color and $\lambda\in(0,1)$.

Step 2. m=m+1, go to Step 1;

Step 3. To normalize p_i of G_b make $\sum_{i=1}^{n} p_i = 1$;

Step 4. Herein, m = n, i.e., the contractive mapping sets between G_b and G_e can correspond one by one. Then according to the value of probabilities which are arranged in the degressive order, we can rearrange the subscript of all the contractive mappings and color mappings in the ascending order for G_b and G_e respectively. After rearranging them, suppose that the new contractive mappings of G_b and G_e is W_k and ψ_k respectively, and the new color mappings of G_b and G_e is B_k and E_k respectively ,where $k=1,2,...,\max(m,n)$.

Step 5. For any given initial point, according to the number of the given in-betweens N and the interpolation function f(t), we can have the middle interpolation formula expression: $M=(1-f(t))G_b+f(t)G_e$, then the contractive mapping sets corresponding to the M are:

$$\eta k\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}\bar{a}_k(1-f(t))+a_k f(t) & \bar{b}_k(1-f(t))+b_k f(t)\\ \bar{c}_k(1-f(t))+c_k f(t) & \bar{d}_k(1-f(t))+d_k f(t)\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}+\begin{bmatrix}\bar{e}_k(1-f(t))+e_k f(t)\\ \bar{f}_k(1-f(t))+f_k f(t)\end{bmatrix}$$

Moreover, the probability and the color are given as $P_k = p_k(1-f(t))+q_k f(t)$, $M_k=(1-f(t))*B_k+f(t)*E_k$ respectively, where $k=1,2,...,\max(m,n)$. Herein, we acquire the united IFS codes, which are used to generate the in-betweens.

Step 6. While i=1, 2, 3… N-1, execute the following two steps repeatedly:

① The in-between M= (1-f (t)) G_b+ f (t) G_e can be generated by IFS, where t=i/N;

② Save the last in-between.

In the following section 6, we discuss and analyze the results of testing our algorithm with the Matlab 6.5 software. Then, we compare the main features between our key frame algorithm and the traditional key frame algorithm.

6 Experimental Result and Analysis of Our Algorithm for Color Fractal Graphics

Here, in this section, we first analyze the experimental result and then summarize the characteristics of our algorithm.

6.1 Analysis of the Result

Since a Barnsley leaf is a classic fractal object, the start key frame (Fig. 1 (a)) we have selected is a Barnsley leaf and the end key frame (Fig. 1 (f)) is also a Barnsley leaf whereas it has been swung by the wind.

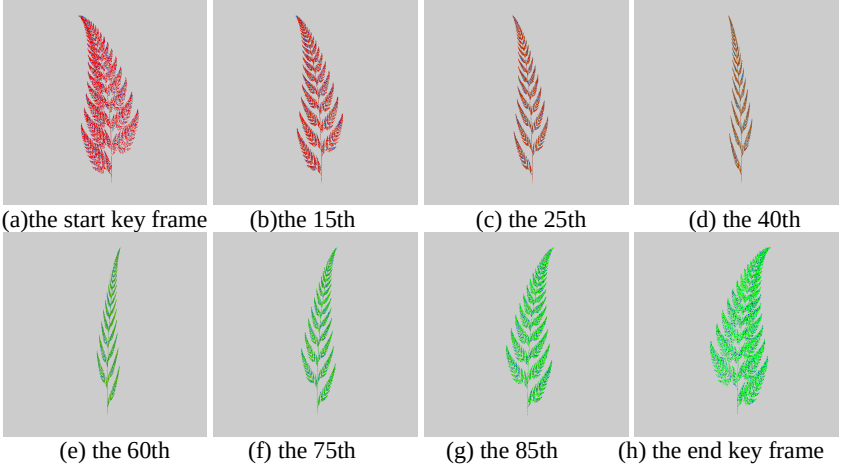


Fig. 1. The generation of the in-betweens

By testing our algorithm with the Matlab 6.5 software which we have used as a test tool, we have succeeded in implementing the algorithm to realize 99 in-betweens between the start and the end key frame on a Pentium IV 2.4 GHz PC. Figure 1 shows inbetweening examples. We have selected only 8 (including Fig. 1 (a) and Fig. 1 (h)) out of 99 in-betweens.

It can be seen from the Figure 1 that the animation generated according to this method has the sense of reality and the beauty of artistic conception which is relatively satisfying. The experimental result shows that each of the in-betweens is a fractal, and that the animation generated according to our method is smooth, natural, and fluent.

6.2 The Characteristics of Our Algorithm

The color objects in nature can be classified into two types. One is the color Euclid objects; the other is the color Fractal objects. In the key frame animation, the former has been widely studied by both national and international researchers [2-8]. However, the latter has been neglected. Just as what we have said before, there are many

color fractal objects in the nature, which the traditional Euclid-based algorithm cannot deal with. Therefore, as a new fractal-based algorithm for the key frame animation, the characteristic of our algorithm is distinctly different from the Euclid-based algorithm. The following are the main differences between two types of the algorithm.

First of all, the study object is different. Our fractal-based algorithm can deal with the fractal object while the traditional Euclid-based algorithm can not do. Since fractal geometry is an alogical architecture based on iterations or recursion, our algorithm is mainly appropriate for the description of the natural objects. But since Euclid geometry is a logical architecture based on axioms, the traditional Euclid-based algorithm is mainly appropriate for the description of the man-made objects.

Secondly, although there are all the three main problems needed to be solved, the essence of the problems is greatly different. For our new fractal-based algorithm, the three problems are: First, the corresponding principles of the contractive mappings. Second, the corresponding principles of the probabilities. Third, the corresponding principles of the colors. But for the traditional Euclid-based algorithm, the three problems are: First, the matching mapping, i.e., which parts or points of the start key frame should be mapped or transformed into which parts or points of the end key frame. Second, the interpolation of the vertexes, i.e., once these correspondences have been established, their trajectories and transformations through time must be determined. Third, the rendering method, i.e., how the in-betweens is to be rendered.

Thirdly, the prerequisite for them is different. The prerequisite for our new fractal-based algorithm is that the acquisition of the IFS codes of an object is a requirement. However, for the traditional Euclid-based algorithm, the prerequisite is that the acquisition of the vertex coordinates or the formulary expression of an object is a requirement.

Fourthly, the dimension number for the fractal object and Euclid object is quite different. A characteristic of a fractal is a mathematical parameter called fractal dimension. Conventional geometry is concerned with regular shapes and whole-number dimensions, such as lines (one-dimensional) and cones (three-dimensional), while fractal geometry deals with shapes found in nature that have non-integer, or fractal dimensions.

Fifthly, the way, by which the point, the line, and the surface are treated, is distinctly different. In our new fractal-based algorithm, they are treated as a whole, but in the Euclid-based algorithm, they are treated as a part.

By summarizing and contrasting our new fractal-based algorithm and the traditional Euclid-based algorithm, it can be seen that our algorithm can deal with the fractal object while the conventional algorithm cannot, and that the fractal idea can be effectively applied in the key frame animation.

7 Conclusions and Future Work

In this paper, a computer aided inbetweening algorithm for color fractal graphics is presented. Our study shows that the fractal idea can be effectively applied in the key frame animation. The main feature of our algorithm is that it can deal with a fractal object while the conventional algorithm cannot. There are three main problem needed to be solved in our algorithm. A relatively good solution to the corresponding principles of the contractive affine transformation, colors, and the probabilities between the

start and the end key frame is put forward. The theoretical and experimental results demonstrate that the animation generated according to our algorithm is smooth, natural, and fluent. In addition, our algorithm has the merits of stability, efficiency, and verisimilitude. In application, our algorithm has many practical values which can improve the efficiency of animation production and simultaneously greatly reduce the cost. However, how to effectively acquire the IFS codes of a fractal object is still a difficult problem deserving further study.

Acknowledgements

The authors wish to acknowledge the support of the National 863 Plan Project (No: 2004AA420100).

References

1. Mandelbrot, B.B.: The Fractal Geometry of Nature. W. H. Freeman and Company, New York (1983)
2. Jin, X.G., Bao, H.J., Peng, Q.S.: A Survey of Computer Animation. Journal of Software, Vol. 8 No. 4 (1997) 241–251 (in Chinese)
3. Seah, H. S., Feng, T.: Computer-Assisted Coloring by Matching Line Drawings. The Visual Computer, Vol. 16 (2000) 269–304
4. Fekete, J. D., Bizouarn, E., Cournarie, E., Galas, T.: A Paperless System for Professional 2D Animation. Computer Graphics Proceedings, Annual Conference Series (1995) 79–90
5. Kort, A.: Computer Aided Inbetweening. Proceedings of the 2nd International Symposium on Non-photorealistic Animation and Rendering, Annecy, France (2002) 125–132
6. Rademacher, P.: View-Dependent Geometry. Proceedings of ACM SIGGRAPH 99 (1999) 439–446
7. Fabian, D.F., Philip, S., Koen, E., Frank, V. R.: Automatic In–Betweening in Computer Assisted Animation by Exploiting 2.5D Modeling Techniques. Proceedings of Computer Animation (2001) 192–200
8. Chen, C.B., Zheng, Y.P.: Study on an Algorithm of Key Frame by Meshing a Three-Dimension Object. Journal of Huazhong University of Science and Technology, Vol. 33 No. 10 (2005) 76–79 (in Chinese)
9. Dang, Y.J.: The Study of the Iterated Function System and Fractal Interpolation. Northwestern Polytechnical University, Xi'an (2000) (in Chinese)

Feature Sensitive Hole Filling with Crest Lines

Mingxi Zhao, Lizhuang Ma, Zhihong Mao, and Zhong Li

Department of computer science and engineering, Shanghai Jiao Tong University
200240, Shanghai, China
{zhaomx, ma-lz, mzh_yu, lizhong}@cs.sjtu.edu.cn

Abstract. Feature sensitive hole filling is important for many computer graphics and geometric modeling applications. In this paper, we address the problem of reconstructing the salient features when filling holes in mesh. It respects fine shape features and works well on various types of shapes, including natural mesh and mechanical parts. For representing salient surface features, we adopt crest lines. The connectivity of the hole including the lost crest lines are decided by triangulation and region growing. The geometric position of the vertices and crest points in the hole are resolved by using a least-squares problem.

1 Introduction

Most shapes are nowadays captured with the range scanning devices, which can give high accuracy and resolution. However, these systems can not obtain a watertight surface, and the resulting models may have holes corresponding to deficiencies in original point data, or to occlusion in the process of scanning. Moreover, certain physics properties of surface such as high reflectivity may also prevent the range scanner from capturing geometry.

In some applications, it is often desirable to fill in these holes because of aesthetics or the need for a watertight mesh from industry. In computer graphics field, there are a lot of methods proposed to fill hole. It can be performed as a post-processing of reconstruction, or it can be integrated into the surface reconstruction algorithm. According to the representative type of the model, hole filling can be classified to two types: mesh-based and point-based. The method proposed in this paper belongs to the former and is performed as a post process. In this subsection, we only give a simple review on two types of methods used as a post-processing after surface reconstruction. Minh X. Nguyen et. al. [1] utilized the geometry images [2] to fulfill the matching both in boundary and context condition with the hole. However, their methods depend on the shape of boundary because of the parameterization, and can not deal with some holes. Furthermore, it can not reconstruct the salient features completely. Liepa [3] firstly use a modified minimum-area triangulation to close the holes, and then the techniques in [4] are used to refine the hole, i.e., adding the vertices and adjusting the shape of the triangle. In order to match the shape, the fairing process is applied. This method produces good-looking result for arbitrary holes in meshes, but it can not reconstruct salient features. Based on this method, we proposed a new method to reconstruct sharp features by using crest lines and the least-squares method.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 660 – 663, 2006.

2 Filling Hole Process

Given an orientable triangle mesh *M* with arbitrary holes, our goal is to fill the hole and there have a natural and nice-looking transition including the salient features between the hole and the existing boundary. Totally, we use crest lines to build salient features, and the topology of the hole and crest lines are built by triangulation and region growing respectively. The geometry of both the vertices and crest points are decided by the least-squares method. In the following, we will give an outline of our proposed method (See Fig. 1).

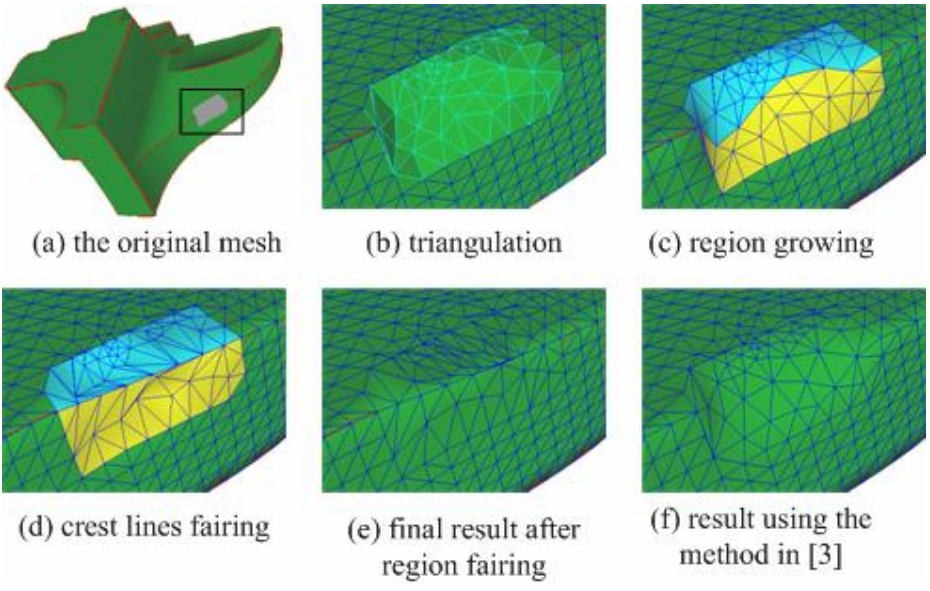


Fig. 1. Overview of our proposed method. In (a), the red lines are the ridges with threshold T = 4 (quantity to measure the strength of a crest line) and K = 2 (the neighborhood size to build the crest lines), the grey region is the hole. (b), (c), (d), (e) and (f) are close-up views of the dark rectangle in (a). (e) and (f) are the final results built by our method and the method in [3], respectively. Obviously, our method can achieve better salient features.

1. **Crest Lines Detection.** The crest lines are salient surface characteristics and are widely used for shape matching, interrogation and segmentation. Given a mesh *M* with the hole, the first step is to detect the ridges and ravines and set proper threshold values to select salient feature lines (Fig.1 (a)). We use the method proposed in [5] to detect the crest lines.
2. **Triangulation and Refinement.** Once a hole is identified (it can be identified automatically by searching boundary vertices or can be indicated by a user interactively), hole triangulation and mesh refinement are applied. Hole triangulation is used to close the hole with a minimum area surface, and in [3] a modified method based on min-max dihedral angle is applied to process arbitrary holes. In order to approximate the density of the surrounding mesh, the patched hole is

refined by relaxing certain edges and subdividing certain triangles. In this step, we use the same method proposed in [3] to build the connectivity of the hole (The cyan edges in Fig.1(b)).

3. **Region Growing.** In order to connect the crest lines correctly in the holes, we perform a region growth according to the surrounding boundary crest points. The result is a segmented region by the identified crest lines (The blue and yellow regions in Fig. 1(c)). Sometimes, in our system the user can also have the ability to connect some crest lines.
4. **Crest lines Fairing.** After region growing, the identified crest lines in the hole have proper topology, but inappropriate geometry. The least-squares method is utilized to decide the geometric position of those crest points (Fig. 1(d)).
5. **Region Fairing and Refinement.** After obtaining proper feature lines in the hole we do mesh fairing to obtain the final geometry of the hole by using the same least-squares method as the crest lines faring. In most situations, mesh refinement is again used to approximate the mesh density because the connectivity has been changed in the fourth step (Fig. 1(e)).

3 Results

Besides the advantages inherited from the method in [3] such as the ability to process arbitrary holes, our proposed method can reconstruct salient features. Fig. 2 shows the best result of our method. There are threes holes in this mechanical part, which has one, two and three ridges passed through them (Fig. 2(a)). The method in [3] can only close the hole (Fig. 2(b)), and the method in [1] got better result but there are some noises in the region of the share edges (Fig. 2(c)), while only our method can reconstruct the sharp features completely (Fig. 2(d)).

When the hole has the ridges and the ravines at the same time, our method can also obtain nice-looking results. In Moai model (Fig 3), the hole has complex boundary that is a challenge job for the parameterization in method [1]. There are two crest

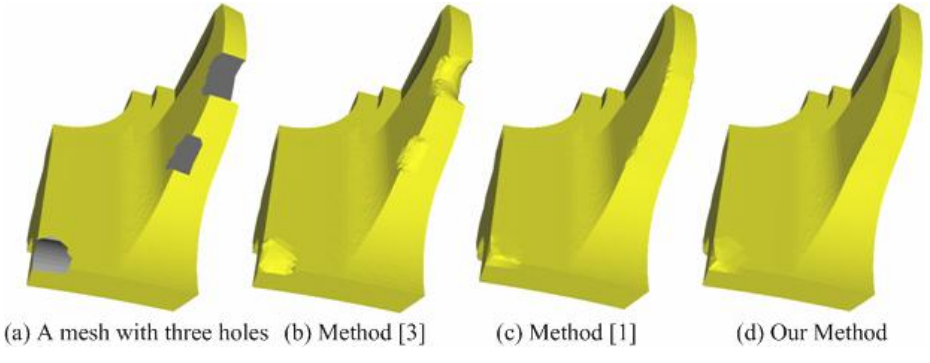


Fig. 2. Compared with other two state-of-the-art methods. It can show that our method achieves the best salient features.

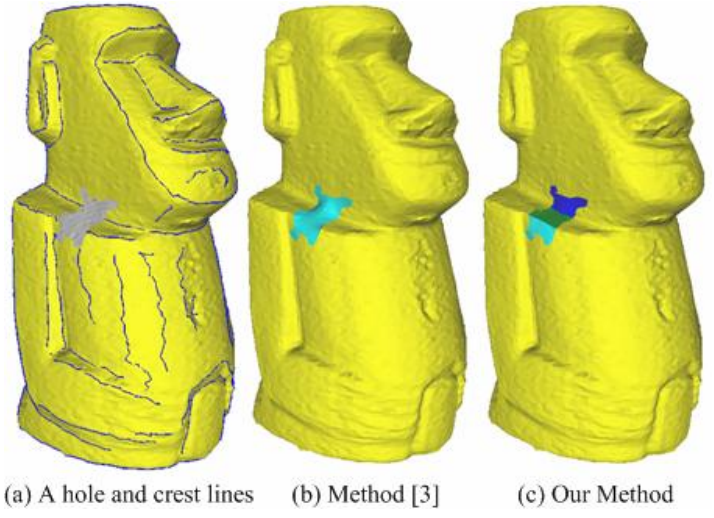


Fig. 3. Compared with the method in [3]. There are ridges and ravines passed through the hole at the same time in (a).

lines, one ridge and the other ravine, passed through the hole(Fig. 3(a)). Although improved by using fairing process, it always gives a minimum-area solution because of lack of considering the salient features (Fig. 3(b)). By using the crest lines in our method, Fig. 3(c) seems more natural.

Acknowledgement. This work was supported in part by national natural science foundation of China (Grant No. 60373070) and postdoctoral research foundation of Shanghai, China (Grant No. 05R214129).

References

1. Minh X. Nguyen, Xiaoru Yuan, Baoquan Chen. Geometry Completion and Detail Generation by Texture Synthesis. In *Proceedings of the Pacific Graphics* 2005
2. Gu, X., Gortler, S.J., Hoppe, H.: Geometry images. In *SIGGRAPH 02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*,pp. 355–361. ACM Press, New York, NY, USA (2002)
3. LIEPA, P. Filling holes in meshes. In *Proceedings of the Eurographics/ACM SIGGRAPH symposium on Geometry processing*, Eurographics Association, 200-205(2003).
4. R. Pfeifle and H.-P. Seidel. Triangular B-Splines for Blending and Filling of Polygonal Holes. In *Proc. Graphics Interface '96*, pages 186-193, 1996.
5. YOSHIZAWA S., BELYAEV A. G., SEIDEL H.-P.: Fast and robust detection of crest lines on meshes. In *Proc. of ACM Symposium on Solid and Physical Modeling* (2005).

A Speech Stream Detection in Adverse Acoustic Environments Based on Cross Correlation Technique

Ru-bo Zhang, Tian Wu, Xue-yao Li, and Dong Xu

College of Computer Science and Technology, Harbin Engineering University, Harbin Heilongjiang, China 150001
zrbzrb@hrbeu.edu.cn

Abstract. Speech signal detection is very important in many areas of speech signal process technology. In real environments, speech signal is usually corrupted by background noise, which greatly affects the performance of speech signal detection system. Correlation analysis is a waveform analysis method which is commonly used in time domain, and the similarity of two signals can be measured by using of the correlation function. This paper presents a new approach based on waveform track from cross correlation coefficients to detect speech signal in adverse acoustic environments. This approach firstly removes irrelevant signal so as to decrease the interference from noise by making use of computing cross correlation coefficients, and then decides whether contains speech signal or not according to the waveform track. Moreover, the performance of the algorithm is compared to the approach based on short-term energy and the approach based on spectrum-entropy in various noise conditions, and algorithm is quantified by using the probability of correct classification. The experiments show that the waveform from cross correlation coefficients is powerful in anti-interference, especially being robust to colored noise such as babble.

1 Introduction

Speech signal detection refers to the ability of distinguishing speech from noise, and is an integral part of a variety of speech signal process systems, such as speech coding, speech recognition and speech enhancement. Meanwhile, the speech signal detection is always pretreated before other speech signal processing system. In the wireless system, for example, the speech detection module is used for discontinuous transmission to save battery power. Moreover, it is a great deal of need to the pure speech detection products. In our life, most of speech signal mixed strong noise is detected by human for the lack of reliable speech detection product. Over a long period of time, the man who works in this kind of environment would injure their hearing ability.

Aiming at the special research field of speech signal detection in adverse acoustic environments, this paper presents a speech signal detection algorithm in adverse acoustic environments based on cross correlation technique. The proposed approach firstly removes irrelevant signal so as to decrease the interference from noise by making use of computing cross correlation coefficients, and then decides whether

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 664 – 667, 2006.

contains speech signal or not according to the waveform track. In next part, we will introduce how to remove the effects from noise signals, and how to extract the characteristics of speech from the waveform track in the proposed algorithm. Finally, the experiments result of the proposed algorithm in white noise and colored noise are given respectively, and a table is also presented to measure the performance of the presented algorithm quantitatively. Compared to the approach based on short-term energy [2] and the approach based on spectrum-entropy in various noise conditions.

2 Speech Signal Detection Based on Cross Correlation Technique

Correlation analysis is a waveform analysis method which is commonly used in time domain. As a correlation analysis, cross-correlation function is used to measure the time lag between the two signals, and also to detect useful signal form noise. If two signals are completely unrelated, the value of cross-correlation function is close to zero. On the contrary, if the two signal waveforms are correlative, the value of each correlative function will be peaks. By using of the characteristic of cross-correlation function, and changing the location of the two signals in the cross-correlation function point by point (mobile length is the length of pitch in time domain.), the waveform track of the cross-correlation function can be gotten . The waveform track produced by this way contains a wealth of information, such as the frequency of pitch and the phase of signal, as well as the frequency and energy of formant. According to the characteristics extracted from the waveform track, a judgment about whether contains speech signal could be made.

3 Judgment by the Waveform Tack of Cross-Correlation Function

The algorithm removes the interference from unrelated signal by the previous signal processing. It is well known that the colored noise is robust to cross-correlation function, and those signals of colored noise are closely correlative. Therefore, the traditional speech detection methods are regularly sensitive to the colored noise. However, this noise exists in many detected signals, and a great deal of equipment in our life trend to produce it. Experiments show that the characteristic of waveform tack introduced by the proposed algorithm is robust to the colored noise.
If the detected signal contains speech signal, its waveform track should accord with the follow characteristics:

Characteristic 1. The maximum of the waveform can not be extremely small. After the process of cross-correlation, the energy of those related signals in speech signal would survive and feedback in the waveform track.
Characteristic 2. The number of those maxima of waveform should range in certain scope. Because the waveform relates directly to the distribution of speech signal frequency, whose highest harmonic generally lower than 5000Hz.

Characteristic 3. The formant of speech looks like a straight line approximately. Therefore, the interval to previous maximum appears in time is approximately equal to next interval, however, It can not absolutely closely coincide to the next interval.
Characteristic 4. There is difference in maxima, some are greater, and some are smaller. It is because of that the greater is produced by all formants of speech signal, and the smaller may be produced by just one formant.
Characteristic 5. If there are three or more consecutive waveform track which coincide to above characteristics, the detected signal is judged to contain speech signal.

4 Experiment and Results

To evaluate the effectiveness of the proposed algorithm, great deals of experiments have been done. The speech database used in the experiments contains the isolated words spoken by 10 male and 10 female speakers from microphone. The noise materials were collected from the NOISEX-92 noise-in-speech database including white noise and factory noise [7]. Furthermore, some sound artifacts were included in our experiments such as breath noise, cough noise and mouse click noise. Each case is a different combination of the speech normalization level, the noise type and the SNR. The SNR value here is computed as the ratio of the total energy of speech to that of the noise over the entire utterance.

An executive program was generated based on the proposed algorithm in the VC 6.0 platform. To evaluate the effectiveness of the proposed algorithm, great deals of experiments have been done. More precise experimental results are summarized in Table 1 compared to the algorithms base on the short-term energy and spectrum entropy. CCWT in Table 1 is the abbreviation for the proposed algorithm.

Table 1. The probability of correct speech detection obtained by short-term energy, spectrum entropy and the proposed algorithm with different types and levels of noise

	SNR	Energy	Entropy	CCWT
White noise	15	0.96	0.97	0.99
	10	0.76	0.78	0.99
	5	0.64	0.70	0.95
	0	0.63	0.56	0.92
	5	0.12	0.30	0.88
Factory noise	15	0.94	0.95	0.97
	10	0.71	0.73	0.94
	5	0.57	0.64	0.92
	0	0.38	0.47	0.87
	5	0.07	0.22	0.81

The results show that the proposed algorithm has an overall better performance than the compared approaches in all the noise types used here. Especially, it performs excellently in factory noise, which is obstinate to most algorithms ordinarily.

5 Conclusion

An algorithm based on waveform track from cross correlation coefficients to detect speech signal is presented in this paper. This approach firstly removes irrelevant signal so as to decrease the interference from noise, making use of computing cross correlation coefficients, and then decides whether contains speech signal or not according to the waveform track. The experiment results show that the proposed approach is robust and accurate to strong noise, and has an overall better performance than short-term energy and spectrum entropy especially in the condition of colored noise. However, our experiments also show that the algorithm is not robust enough to very strong noise with SNR under -5dB or more.

Acknowledgements

This study was supported by the National Nature Science Foundation of China (No.60475016).

References

1. Jia, C., Xu, B.: An Improved Entropy--based Endpoint Detection Algorithm. International Conference on Spoken Language Processing (ICSLP 2002), Taipei (2002) 285–288
2. Rabiner, L., Juang, B.H.: Fundamentals of Speech Recognition. Prentice Hall PTR, New Jersey (1993)
3. Bullington, K., Fraser, I.M.: Engineering aspects of TASI. Bell System Technical Journal, 38 (1959) 353–364
4. Zhu, S.Q., Qiu, X.H.: Research on Endpoint Detection of Speech Signals, Computer Simulation, 22 (2005) 214–216
5. Chen, L., Zhang, X.W.: New Methods of Speech Segmentation and Enhancement Based on Fractal Dimension. Signal Processing Proceedings, (2000) 281–284.
6. Julien, P., Rouas. Jean-Luc, Régine, A.O.: Robust speech / music classification in audio documents. Dans : International Conference on Spoken Language Processing (ICSLP'2002), Denver 3 (2002) 2005–2008.
7. Varga, A.P., Steeneken, H.J.M., Tomlinson, M., Jones, D.: The NOISEX-92 Study on the Effect of Additive Noise on Automatic Speech Recognition. DRA Speech Research Unit Technical Report, (1992)

Contour Construction Based on Adaptive Grids

Jinfeng Yang, Renbiao Wu, Ruihui Zhu, and Yanjun Li

Tianjin Key Lab for Advanced Signal Processing
Civil Aviation University of China, P.O. Box 9, Tianjin, P.R. China
{jfyang, rbwu}@cauc.edu.cn

Abstract. Contour information is always valuable for object analysis in image processing. In this paper, a new method of constructing contours of skin regions is proposed. To exploit skin formation in images, a nonlinear skin color classifier is first introduced. Then, a region splitting scheme is adopted to generate adaptive grids over skin regions. Based on the grids, initial contours are constructed. Finally, the contours are refined according to the minimum energy principle. Experimental results show that the proposed method has a good performance in contour construction of skin regions.

Keywords: Skin detection, region splitting, contour construction.

1 Introduction

The contour information is very useful in object tracking, shape analysis and image segmentation and so on. Traditionally, the edge-based and region-based segmentation techniques are often used for contour acquisition. The common disadvantage of these methods is that they often fail to extract valuable contours when the environments of images are cluttered. Since the introduction of "snake" based on the theory of curve evolution[1], extensive studies have been done on active contour models (ACMs) that can drive an initial closed curve towards the edges of regions of interest (ROIs) [2], [3], [4], [5], [17].

However, the application of the active contour models is often limited by two problems. The first is the sensitivity to initialization. The ACMs without manual initialization often fail to construct reliable object contours in the evolution process. Moreover, the initial shape of a closed curve also affects the evolving results directly. Thus, the ACMs can not meet the requirements of obtaining object contours automatically in practice. The other is the computational cost. The evolutionary process of driving a closed curve towards the boundary of an object is often time-consuming under a certain energy criterion, especially involving region functionals. This makes the ACMs unsuitable for extracting object contours in some applications, such as shape analysis [6] and gait recognition [7].

In this paper, we have no intention of addressing the general approaches to extracting object contours. Instead, we concentrate on a specific method for effectively obtaining the contours of skin regions. This is very useful for skin segmentation, face detection and pornographic image detection, because the contours of skin regions can greatly facilitate the localization of the potential skin

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 668–678, 2006.

regions, the process of face detection [8] and the discrimination between faces and naked bodies [9].

To start with, we briefly introduce our method of building a nonlinear skin color classifier. Next, the proposed method is detailed that can automatically and efficiently accomplish the contour construction of skin regions. Experimental results show that our proposed method performs well in obtaining the contours of skin regions.

2 Nonlinear Skin Color Classifier

The human skin colors cluster spatially in whatever color space used [10]. However, an accurate description of skin color distribution is still not achieved at present [11]. Here, we present a new nonlinear space transformation for investigating the skin and non-skin color distributions. Fig. 1 shows our main idea to skin and non-skin color discrimination.

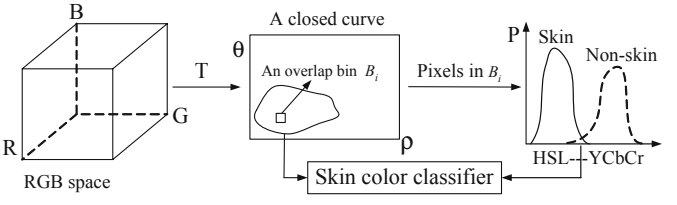


Fig. 1. The framework of the proposed a nonlinear skin color classifier

The proposed nonlinear space transformation is defined as

$$\begin{cases} \rho = ((\alpha_1 R - \alpha_2 G)^2 + (\alpha_2 G - \alpha_3 B)^2)^{1/2} \\ \theta = \begin{cases} \theta_{t_1}, & \alpha_1 R - \alpha_2 G \geq 0 \\ \pi + \theta_{t_2}, & \alpha_1 R - \alpha_2 G < 0, \alpha_2 G - \alpha_3 B \leq 0 \\ 2\pi - \theta_{t_2}, & \alpha_1 R - \alpha_2 G < 0, \alpha_2 G - \alpha_3 B > 0 \end{cases} \end{cases} \tag{1}$$

where $\theta_{t_1} = \arccos((\alpha_2 G - \alpha_3 B)/\rho), \theta_{t_2} = \arccos(|\alpha_2 G - \alpha_3 B|/\rho)$, and $\alpha_1, \alpha_2, \alpha_3$ represent the normalized r, g, b respectively. To obtain the skin and non-skin distributions, we manually build a pixel dataset that contains 40,803,483 skin samples and 54,710,462 non-skin samples using the ECU face image database [15], and construct the skin and non-skin color histograms in (ρ, θ) plane(see Fig. 2(a)). From Fig. 2(a), we can see that the skin colors prominently cluster in the proposed space, while the non-skin colors spread out along the θ axis.

Based on the statistical results, the skin color distribution is fit using a closed curve (see Fig. 2(b)), which enclose 98.4% of skin pixels in our skin set. The closed curve is expressed as

$$S(\rho, \theta) = \begin{cases} b^2(\tilde{\rho}_1 - \rho_0)^2 + a_1^2(\tilde{\theta}_1 - \theta_0)^2 - a_1^2 b^2, f_1 \leq 0 \\ b^2(\tilde{\rho}_2 - \rho_0)^2 + a_2^2(\tilde{\theta}_2 - \theta_0)^2 - a_2^2 b^2, f_2 \leq 0 \\ b^2(\tilde{\rho}_3 - \rho_0)^2 + a_3^2(\tilde{\theta}_3 - \theta_0)^2 - a_3^2 b^2, f_3 \geq 0 \end{cases} \tag{2}$$

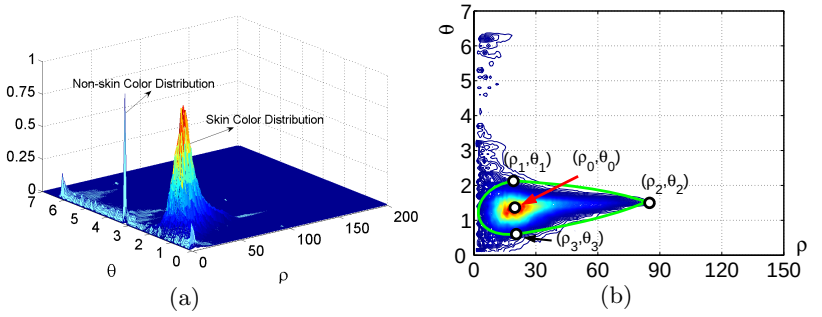


Fig. 2. Statistical result of pixels. (a) The skin and non-skin distributions. (b) A closed curve enclosing the skin pixels.

Table 1. The values of some parameters

(ρ, θ)	0	1	2	3	$\beta_1(a_1)$	$\beta_2(a_2)$	$\beta_3(a_3)$
ρ_-	20.2	19.3	85	21.1	$\pi/36$	$-\pi/36$	$\pi/36$
θ_-	1.37	2.15	1.45	0.62	(20.2)	(64.2)	(64.2)

where $f_1 = \rho - \rho_0, f_2 = (\rho^2 - \rho(\rho_1+\rho_2)+\rho_1\rho_2)(\theta-\theta_2), f_3 = (\rho^2-\rho(\rho_2+\rho_3)+\rho_2\rho_3)(\theta-\theta_2), \tilde{\rho}_i = \rho\cos\beta_i - \theta\sin\beta_i, \tilde{\theta}_i = \rho\sin\beta_i + \theta\cos\beta_i$. The three sub-curves constrained by f_1, f_2 and f_3 are segmented piecewise from three ellipses locating at the same center (ρ_0, θ_0). β_i denotes the rotation angle of the ith sub-curve in the (ρ, θ) plane. The parameters in Eq. (2) are obtained empirically and listed in Table 1.

Suppose that the skin and non-skin pixels overlapping in the ith $\rho_i\theta_i$ bin build two sets ω_{1i} and ω_{2i}, and $P(\omega_{1i})$ and $P(\omega_{2i})$ are their prior probabilities, respectively. Assume $d_1 = S/L$ and $d_2 = Cb/Cr$, where S, L, Cb and Cr are color components in HSL and YCbCr color spaces, we define $x = \sqrt{d^2 + \alpha^2}$, where $d = \sqrt{d_1^2 + d_2^2}$ and $\alpha = \arccos(d_1/d)$. Hence, two 1D histograms with 0.1 bin width in x are populated by the pixels in ω_{1i} and ω_{2i}. Based on the Bayes decision rule for minimum error, we can build a discriminant function

$$g(x) = p(x|\omega_{1i})P(\omega_{1i}) - p(x|\omega_{2i})P(\omega_{2i}) \tag{3}$$

Note that we let $g(x) = 1$ for the pixels that are not in the overlap bins enclosed by the closed curve $S(\rho, \theta)$. Therefore, the non-linear skin color classifier is defined as

$$F_s(q) = S(\rho_q, \theta_q)g(x_q) \tag{4}$$

where q is a pixel. If $F_s(q) \leq 0$ is satisfied, it is labelled as skin.

3 Contours of Skin Regions

3.1 Adaptive Grid over Skin Regions

The mechanism of region splitting is a fundamental processing stage in our contour construction. Here, a region splitting scheme is proposed that can generate

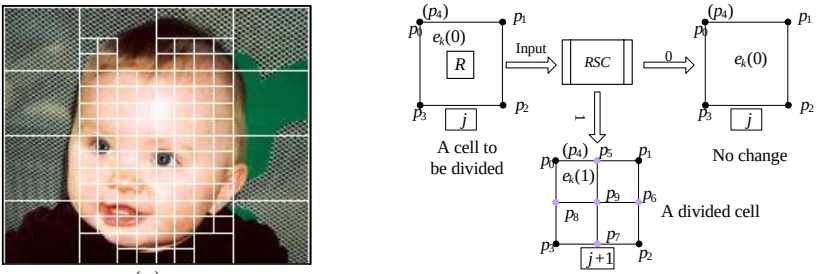


Fig. 3. Grid generation. (a) Grids over skin regions. (b) Region splitting scheme.

adaptive grids over ROIs automatically under our defined region splitting criterion (RSC), as shown in Fig. 3(a).

Assume that e_k denote the kth cell of a grid at the jth splitting iteration level, $e_k(0)$ and $e_k(1)$ denote two states of the cell, the nodes of the cell in the two states are numbered as shown in Fig. 3(b). Thus, the node relationship between two adjacent splitting levels can be described as

$$\begin{cases} p_i^{e_k(1)} = p_i^{e_k(0)}, & i = 0 \sim 4 \\ p_{i+5}^{e_k(1)} = (p_i^{e_k(0)} + p_{i+1}^{e_k(0)})/2, & i = 0 \sim 3 \\ p_{i+5}^{e_k(1)} = (p_i^{e_k(0)} + p_{i-2}^{e_k(0)})/2, & i = 4 \end{cases} \tag{5}$$

The skin color and texture information contained in R are here exploited together to build a region splitting criterion (RSC). For reducing the computational cost of generating grids over regions, we only select a square region R occupying one ninth of a cell as its representative (see the top-left of Fig. 3(b)).

Suppose that U_1 and U_2 are the set of all pixels and the set of all detected skin pixels in R respectively, we can calculate the approximate probability density of occurrence of the intensity levels in the two sets: $I_k(\tau) = h_k(\tau)/N_k(\tau = 0, 1, ..., G_k - 1; k = 1, 2)$, where N_k is the total numbers of pixels, G_k is intensity levels in U_k, and $h_k(\tau)$ is the number of pixels in the τth intensity level. Thus, the following statistics can be obtain [12]

$$\begin{cases} \mu_k = \sum_{\tau=0}^{G_k-1} \tau I_k(\tau), \sigma_k^2 = \sum_{\tau=0}^{G_k-1} (\tau - \mu_k)^2 I_k(\tau) \\ \mu_{4k} = \sigma_k^{-4} \sum_{\tau=0}^{G_k-1} (\tau - \mu_k)^4 I_k(\tau)^{-3} \end{cases} \tag{6}$$

where $\mu_{4k}(k = 1, 2)$ is a measure of flatness of the histogram. Let $\zeta_1 = A_{skin}/A_R$ and $\zeta_2 = \mu_{42}/\mu_{41}$, where A_{skin} denotes the detected skin area in R, and A_R represents the area of R, the proposed RSC is defined as

$$RSC(R) = \begin{cases} 1, & \text{if } \zeta_1 \geq \varsigma_1 \text{ and } \zeta_2 \geq \varsigma_2 \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

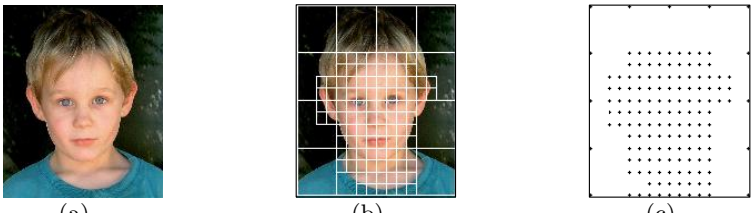
where $\varsigma_1 = 0.35$ and $1 \leq \varsigma_2 \leq 1.6$. The introduction of ζ_2 aims to neglect the regions with high textures considering the homogeneity of skin regions. Thus, a

grid cell is split into four parts equally if $RSC(R) = 1$ is satisfied, otherwise it stays unchanged, as shown in Fig. 4(b).

Let $P_{new}^{e_k(1)} = \{p_5^{e_k(1)}, \cdots, p_9^{e_k(1)}\}$ and $P_j^{e_k(0)} = \{p_0^{e_k(0)}, \cdots, p_4^{e_k(0)}\}$, where $p_0^{e_k(0)} = p_4^{e_k(0)}$, the node relationship between adjacent levels of a split cell is

$$\begin{cases} P_{j+1}^{e_k(1)} = P_j^{e_k(0)}, & \text{if } RSC(R^{e_k(0)}) = 0 \\ P_{j+1}^{e_k(1)} = P_j^{e_k(0)} + P_{new}^{e_k(1)}, & \text{if } RSC(R^{e_k(0)}) = 0 \end{cases} \tag{8}$$

Let $Q_{j+1} = P_{j+1}^{e_0(1)} \cup \cdots \cup P_{j+1}^{e_N(1)}$, where N is the number of cells of a grid at the $(j+1)$th level, we can obtain a node set Q_M after a grid is updated M times. Eliminating the nodes with identical positions in Q_M, we can obtain a node set $Q^*(\subset Q_M)$, as shown in Fig. 4(c).



(a) (b) (c)

Fig. 4. Region split result. (a) An original image. (b)The grid on skin region. (c) The set Q^* from the grid.

3.2 Point Filter and Initial Contours

From Fig. 4(c), we can see that not every node (called "points" in the following) is essential for contour construction, so the redundant points in Q^* should be eliminated in practice. Here, the low-level information within a block centered at a point p_i is also used to describe its property, as shown in Fig. 5(a). The size of a block is $\frac{2}{3}\delta_x \times \frac{2}{3}\delta_y$, where δ_x and δ_y denote the width and the height of a small cell respectively. Assume that s_i denotes the ith block, we define

$$ID(p_i) = \begin{cases} 1, & \text{if } RSC(s_i) = 1 \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

Let E be a set of points with skin information, for $\forall p_i \in Q^*$, if $ID(p_i) = 1$ is satisfied, then $p_i \in E$. One example is shown in Fig. 5(b). For the interior points in E, we propose a point set filter (PSF) to filter them since they always have negative effects on contour construction based on computational geometry rules. In general, an image may contain several skin regions, so E may consist of several subsets. Without loss of generality, assume that $Z_n(\subset E)$ denotes the nth set of points with connectivity in the grid and B_n is a "purified" set corresponding to Z_n. For $\forall p(x, y) \in Z_n$, we can obtain its 4-neighbor points, namely $p_1(x - \delta_x, y), p_2(x, y - \delta_y), p_3(x + \delta_x, y)$ and $p_4(x, y + \delta_y)$, in the grid. The proposed PSF is defined as

$$PSF(p) = \begin{cases} 1, & \text{if } 0 < \sum_{i=1}^{4} ID(p_i) < 4 \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

Thus, if $PSF(p) = 1$ is satisfied, then $p \in B_n$. Fig. 5(c) is the result of PSF applied to Fig. 5(b). To obtain a reliable contour of a set of points, such as B_n, we here modify Jarvis's March that is a classical algorithm with high efficiency in computing the convex hull for T points [13]. Given two vertexes p_{m-1} and p_m, the criterion used in modified Jarvis's March (MJM) for finding the next vertex p_{m+1} is defined as

$$D_{m+1} = \arg\min_{i \in H}(D_i), \begin{cases} D_i = \sqrt{\alpha_i^2 + \ell_i^2} \\ \ell_i = \|p_i - p_m\| \end{cases} \tag{11}$$

where α_i is a anti-clockwise angle between two segments: $\overline{p_{m-1}p_m}$ and $\overline{p_m p_i}$, and H represents the interval (m, T). Fig. 5(d) shows the result of MJM, in which we place a small square over each point satisfying the criterion.

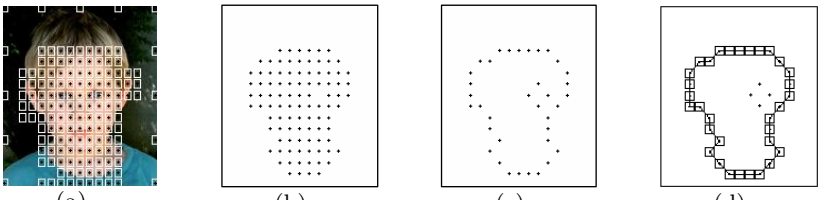


Fig. 5. Initial contour construction. (a) Blocks centering at points. (b) The points with skin information. (c) The set of points after PSF. (d) An initial contour.

3.3 Contour Refinement

To refine the initial contours, we here use the minimum energy principle under a virtual gravity field. Assume that p_1, p_2 and p_3 are three adjacent points on an initial contour, and $\mathbf{\Gamma}$ denotes the exact boundary, their spatial relations are shown in Fig. 6(a).

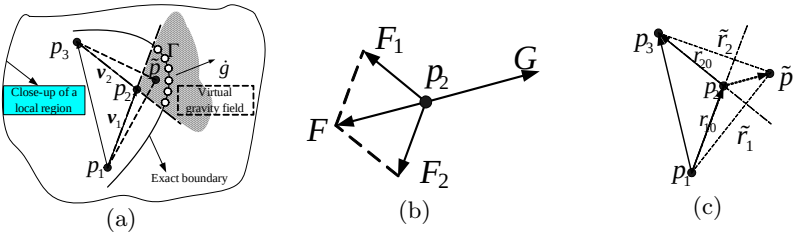


Fig. 6. The state of p_2 in virtual gravity field. (a) Virtual gravity environment. (b) Forces acting on p_2. (c) Point movement.

In Fig. 6(a), p_1 and p_3 are viewed as two fixed points, and p_2 is the point to be adjusted. The potential positions of p_2 are confined in the radial area determined by the supplementary angle between $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$ (see shadow region in Fig. 6(a)). To describe the state of p_2, we refer to p_1, p_2 and p_3 as three particles with unit weight in a virtual gravity field denoted by $\dot{g}$. Based on the gravitational theory, $f_1 = K/r_1^2$ and $f_2 = K/r_2^2$ represent the gravitational forces between p_1 and p_2 and between p_2 and p_3, respectively, where K is the gravitational constant and $r_i (i = 1, 2)$ denotes the Eulerian distance between two particles. Assume that the particle p_2 is unstable due to the action of forces on it, as shown in Fig. 6(b), where $\boldsymbol{G}$ denotes the gravitational force caused by $\dot{g}$, and $\boldsymbol{F}$, less than $\boldsymbol{G}$ in module, is the composition of $\boldsymbol{F_1}$ and $\boldsymbol{F_2}$. Hence, p_2 can move in the direction denoted by $\boldsymbol{G}$, as shown in Fig. 6(c), where denotes a new position of p_2 after movement. When p_2 arrives at $\tilde{p}$ (see Fig. 6(c)), the total consumed energy is

$$W(\tilde{p}) = K(\int_{r_{10}}^{\tilde{r}_1} r_1^{-2} dr_1 + \int_{r_{20}}^{\tilde{r}_2} r_2^{-2} dr_2) = K(\frac{1}{r_{10}} - \frac{1}{\tilde{r}_1}) + K(\frac{1}{r_{20}} - \frac{1}{\tilde{r}_2}) \quad (12)$$

where $\tilde{r}_1 = ||\tilde{p} - p_1||, \tilde{r}_2 = ||\tilde{p} - p_3||, r_{10} = ||p_2 - p_1|| r_{20} = ||p_2 - p_3||$, K is viewed as one, and $||\cdot||$ denotes the Euclidean norm. Based on the minimum energy principle, an object in unstable state always reaches its stable position at the least cost of its energy. Assume that $\boldsymbol{\Gamma}$ represents the balanced boundary of all unstable points, the position with minimum $W(-)$ on $\boldsymbol{\Gamma}$ can be regarded as the optimal balanced position of p_2. Hence, in the motion area of p_2 indicated in Fig. 6(a), only the positions on the skin edges are valuable. So, to detect the real boundaries of skin regions, we first define a skin edge detector as

$$\Psi(x, y) = \begin{cases} 1, & \sum_{i=0}^{4} \xi_i \neq 4 \text{ for } \xi_0 = 1 \\ 0, & \text{otherwise} \end{cases}, \quad \xi_i = \begin{cases} 1, & F_s(q_i) \leq 0 \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

where q_0 is the center pixel of its 4-neighbors. Then we combine it with the Canny operator in 8-neighbors considering edge noise suppression in images. The combined edge detector is defined as

$$Edge(x_q, y_q) = \begin{cases} 1, & \eta \neq 0 \\ 0, & \text{otherwise} \end{cases}, \quad \eta = \sum_{i=0}^{8} (\Psi(x_{q_i}, y_{q_i}) \otimes C(x_{q_i}, y_{q_i})) \quad (14)$$

where q is the center pixel of its 8-neighbors, $C(x, y)$ represents the Canny operator and $\otimes$ denotes "AND" operator. Hence, the object function for contour refinement is defined as

$$p_i' = \arg\min_{q \in \Omega}(W(q)) \quad (15)$$

where $\Omega = \{q | Edge(x_q, y_q) = 1\}$ is the set of points satisfied to Eq. (14). Thus, for all the points on an initial contour, if their positions with the minimum energies are determined in the counterclockwise order, the initial contour can be refined accordingly.

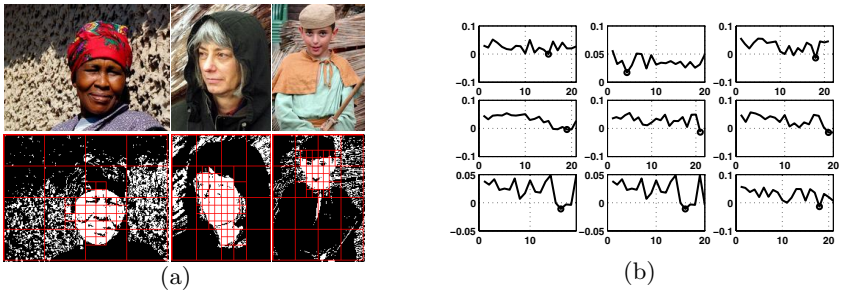
4 Experiments

4.1 Skin Color Detection

We here use the Compaq image database [14] to test the performance of the proposed skin color classifier. The database consists of 4670 skin images and 4997 non-skin images. Based on the image database, we construct two pixel sets Q_1 and Q_2 as testing sets. Q_1 consists of 62,834,397 skin pixels, and Q_2 of 195,438,297 non-skin pixels. Two pixel sets from ECU face database [15] are used as training sets. Here, Jones-Rehg's method [14] is tested at a threshold 0.4 and the results of ECU are reported by its "SkinColorDetector", which can be found in [15]. The CCRs (Correct Classification Rate) and FARs (False Accept Rate) are illustrated in Table 2, which shows that our proposed method can reduce the false positive rate significantly as well as increase the true positive rate.

Table 2. Skin and non-skin pixel detection

Methods	Training set		Testing set	
	CCR(%)	FAR(%)	CCR(%)	FAR(%)
Jones-Rehg	90.88	24.31	90.14	24.91
ECU	93.26	20.78	94.09	21.84
Proposed	**95.46**	**12.34**	**95.03**	**13.01**



(a) (b)

Fig. 7. Adaptive grids and energies of particles. (a) Grids over some images with textures. (b) The energies of some unstable particles.

4.2 Grids over Skin Regions

Based on the above discussions, we can see that the proposed skin color classifier still is a pixel-based method. Hence, false classification is unavoidable, and unexpected noises are always generated in images when the colors of background are similar to human skin's, as shown in Fig. 7(a). To eliminate the possible negative effect of pixel-based skin detection on the generation of adaptive grids, we introduced texture features of a representative region in RSC. Fig. 7(a) shows that the divided cells are not placed over the noisy skin regions but real skin

regions. This shows that the proposed region splitting scheme can successfully find homogenous skin regions while neglecting the regions with high texture information. So, the proposed RSC can greatly reduce the negative effects of noisy skin detection on contour construction of skin regions.

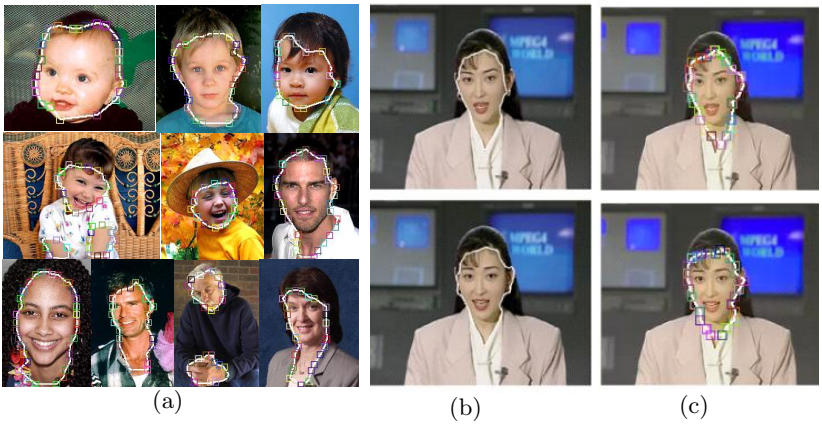
4.3 Energies of Unstable Particles

To understand the energy changes of unstable particles intuitively , the energy of some particles corresponding to the face in the middle column of Fig. 7(a) are illustrated in Fig. 7(b). From Fig. 7(b), we can see that, to any unstable particle, the optimum position can always be obtained that makes the particle's energy minimum in its activity area. Hence, the minimum energy principle is suitable for refining the initial contours of skin regions.

4.4 Contour Construction and Comparison

The ECU face database here is used to test the performance of the proposed method in contour construction. The process of contour construction is accomplished without manual initialization, and suited for obtaining multiple skin region contours. From Fig. 8(a), we also can see that the proposed method is able to successfully describe the skin regions when the backgrounds and the skin regions are very similar in colors. Hence, the fusion of skin color and texture information is essential for detecting real skin regions, and Eq. (15) has a good performance in determining optimal positions for the points to be adjusted in their activity areas. From the results, we can see that the proposed method perform well in obtaining the contours of skin regions.

To compare the proposed method with the ACMs in object contour acquisition, we choose the **DREAM^{2}S** method [17], which has a high ability in describing the contours of regions with homogeneous property. Using manual



(a) (b) (c)

Fig. 8. Contour construction method comparison. (a) Contours of skin regions. (b) **DREAM^{2}S** method with iteration 1400. (c) The proposed method with iteration 1.

initialization with a circle in the skin regions, the **DREAM^{2}S** gives the results shown in Fig. 8(b) at the cost of 1400 iterations. Compared to this, our proposed method without manual initialization gives the contour construction results only after one times iteration, as shown in Fig. 8(c). Hence, from the result shown in Fig. 8(b) and Fig. 8(c), we can see that the proposed method is comparable to **DREAM^{2}S** as well as saving manual work and computational time significantly.

5 Conclusions

We have introduced a new method of contour construction for skin regions. First, a nonlinear skin color classifier has been built to skin color detection. Second, a region splitting scheme has been designed to generate adaptive grids on images. Based on the grids, the initial contours were constructed. Finally, the contour refinement were finished using the minimum energy principle. Experimental results have shown that the proposed method performed well in contour construction of skin regions.

Acknowledgements

This work is supported by NSFC (Grant No. 60325102, 60428101), the open project of CAUC and the start projects of CAUC (Grand No. 05qd02q, 05yk22m).

References

1. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. Int. J. of Computer Vision. **1** (1998) 321-332
2. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. Int. J. of Computer Vision. **22(1)**(1997) 61-79
3. Xu, C.Y., Prince, J.L.: Generalized gradient vector flow external forces for active contours. Signal Processing. **71** (1998) 131-139
4. Chesnaud, C., Réfrégier, P., Boult, V.: Statistical region snake-based segmentation adapted to different physical noise models. IEEE Transactions on Pattern Analysis and Machine Intelligence. **21** (1999) 1145-1156
5. Paragios, N., Deriche, R.: Geodesic active regions: a new framework to deal with frame partition problems in computer vision. Journal of Visual Communication and Image Representation. **13** (2002) 249-268
6. Loncaric, S.: A survey of shape analysis techniques. Pattern Recognition. **31(8)** (1998) 983-1001
7. Collins, R., Gross, R., Shi, J.: Silhouette-based human identification from body shape and gait. In Proc. of Int. Conf. on Automatic Face and Gesture Recognition. (2002) 366-371
8. Yang, M.H., Kriegman, D., Ahuja, N.: Detecting faces in images: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. **24(1)** (2002) 34-58
9. Forsyth, D.A., Fleck, M.M.: Automatic detection of human nudes. Int. J. of Computer Vision. **32(1)** (1999) 63-77

10. Vezhnevets, V., Sazonov, V., Andreeva, A.: A survey on pixel-based skin color detection techniques. Graphicon-2003, Moscow, Russia(2003)
11. Shin, M.C., Chang, K.I., Tsap, L.V.: Does colorspace transformation make any difference on skin detection?. In Proc. IEEE Workshop on Application of Computer Vision(2002)
12. Materka, A., Strzelecki, M.: Texture analysis methods - A review. COST B11 report. Brussels(1998)
13. Jarvis, R.A.: On the identification of the convex hull of a finite set of points in the plane. Information Processing Letters. **2** (1973) 18-21
14. Jones, M.J., Rehg, J.: Compaq skin database. http://www.crl.research.digital.-com/publications/techreports/abstracts/98_11.html
15. Phung, S.L.: ECU face detection database. http://www.soem.ecu.edu.au/-∼sphung/face_detection/database/
16. Jones, M.J., Rehg, J.: Statistical color models with application to skin detection. Int. Conf. on Computer Vision and Pattern Recognition(1999)
17. Jehan-besson, S., Barlaud, M.: $DREAM^2S$: Deformable regions driven by an Eulerian accurate minimization method for image and video segmentation. Int. J. of Computer Vision **53(1)** (2003) 45-70

e-Shadow: A Real-Time Avatar for Casual Environment

Yangmi Lim[1] and Jinwan Park[2]

[1] School of Media and Information Sungshin Women's University
249-1, Dongseon-dong 3-ga, Seongbuk-gu, Seoul, Korea
ymlim@sungshin.ac.kr
[2] Graduate School of Advanced Imaging Science, Multimedia & Film, Chung-Ang University 221, Heukseok-dong, Dongjak-gu, Seoul, Korea
jinpark@cau.ac.kr

Abstract. In this paper, we present a realistic avatar called e-Shadow. This avatar is a virtual shadow of a user and facial animations are injected into the head position as response to the user's gestures. We adopt background subtraction method to separate the user's shape from the background scene and combine several image processing methods in order to reduce noise. A heuristic algorithm for tracking positions of head and hands is proposed to detect predefined user's gestures. Because e-Shadow assumes everyday indoor lighting, casual camera, and small desktop, it can be directly incorporated with video chatting programs as a realistic avatar and the presented detection method might be applied to the applications of real-time interactive media art.

1 Introduction

This paper presents a real-time avatar called e-Shadow as a preliminary but practical application of gesture-based interaction through the Internet. e-Shadow has different characteristics from the existing avatars in that the motion of the avatar is a virtual shadow drawn from the real motion of a user and facial expression of the avatar is changed in real-time as response to user's gestures.

In the segmentation stage for separating user's shape from the original scene, we adopted a background subtraction method [1]. To tackle the obstacles of this stage, noise of diffusing lighting and the effect of shadow, we applied several image processing methods: (1) a convolution technique, a kind of smoothing technique; (2) dilation and erosion; (3) a labeling procedure to eliminate the isolated regions; (4) a sweeping process for filling in gaps between pixels.

For the tracking stage, we devised a heuristic algorithm which tracks the positions of body segments, head and hands. Because we confine our work to the detection of head position and several simple predefined hand gestures expressed by positions and directions of hands, the algorithm focuses on finding distinguishable convex contours by examining surrounding pixels of each pixel in the body segment.

We conjecture that e-Shadow can be used in web communication such as video chatting as a realistic avatar. In addition, the proposed detection scheme of e-Shadow can be applied to the research areas of interactive media art.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 679 – 682, 2006.

2 Segmentation of Virtual Shadow

To separate the user's shape from background, we use a background subtraction method. With this method, the image of the user is subtracted from a previous image without the user. However, the method is extremely sensitive to changes of scenes due to lighting and extraneous events. To solve this problem, we use the background primal sketch [2]. The background primal sketch B_{jk} is calculated by:

$$B_{jk} = median(I^1_{jk}, I^2_{jk}, ..., I^N_{jk}), \tag{1}$$

where $I^1_{jk}, I^2_{jk}, ..., I^N_{jk}$ is a sequence of N recent images, and (j, k) indicates the pixel location. The difference image is taken from a scene which the user is in, after the background primal sketch and a threshold value are determined. In order to decide the pixels which form the difference image, we define d_{jk} as follows:

$$\begin{aligned} d_{jk} &= 1, \text{ if } |I^i_{jk} - B_{jk}| > T \\ &\quad 0, \text{ otherwise} \end{aligned} \tag{2}$$

where I^i_{jk} is the incoming image, B_{jk} is the primal sketch, and T is the threshold.

After the difference picture is obtained, several noise reduction techniques are then performed. First, a convolution technique using a median mask, which is a kind of smoothing technique, is applied to remove Gaussian noise and dilations and erosions are repeated to obtain a clearer outline by removing stray pixels near to the outline. Next, isolated regions outside the shadow are eliminated using an algorithm identifying each cluster of adjacent pixels with the foreground color [3]. Finally, sweeping process removes large gaps intervening between pixel strings in each row.

3 Tracking Head and Hand Shape

We present a new feature-based tracking algorithm [4] tailored for our constrained setting. The idea of this algorithm is to identify center positions of body parts by just analyzing several surrounding pixels. Because we are to detect predefined active gestures, it may assume that all the areas of interest to be detected have convex contours. Under this assumption, the algorithm tries to find pixels whose surrounding pixels within some small distance have the background color and regards them as center positions. Following is the outline of the algorithm. First, we find the center position (x, y) of head, satisfying:

$$\begin{aligned} &\text{color of } (x + o_x, y + o_y) \text{ is white, for } (o_x, o_y) \in O - \{(0, -\alpha), (0, 0)\} \text{ and} \\ &\text{color of } (x + o_x, y + o_y) \text{ is black, for } (o_x, o_y) \in I \end{aligned} \tag{3}$$

where O={ $(x, y) \mid x = -\alpha, 0, \alpha$ and $y = -\alpha, 0, \alpha$}, I = { $(x, y) \mid x = -\beta, 0, \beta$ and $y = -\beta, 0, \beta$}, $\beta = k\alpha > 0$ and $0 < k < 1$. The process of finding the center starts with some large α and iteratively update α to a smaller value until a pixel satisfying above condition is found. The left of figure 1 shows an example of a center position found. Pixels on the outer rectangle, except the one which is nearest to the center of body, are white. Pixels on the inner rectangle are black. The value of α finally used for center detection

varies with the distance between the camera and the user. After the center of head is found, the run length of black pixels including the pixel of the center, l , is calculated for later use in the next step.

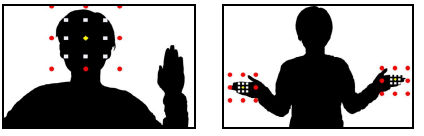


Fig. 1. An example of detected centers of convex shapes

To detect hand gestures, we find center positions of upward, leftward, or rightward hands. For this, a condition similar to (3) with the same O and I but with a different initial value of α scaled by l is used:

$$\begin{aligned} &\text{color of } (x + o_x, y + o_y) \text{ is white, for } (o_x, o_y) \in O - S \text{ and} \\ &\text{color of } (x + o_x, y + o_y) \text{ is black, for } (o_x, o_y) \in I, \end{aligned} \tag{4}$$

where S is $\{(0, -\alpha), (0, 0)\}$, $\{(0, 0), (\alpha, 0)\}$, and $\{(-\alpha, 0), (0, 0)\}$ for an upward, leftward, and rightward hand respectively (see the right of fig.1). Each time a hand position is detected, it is checked if there is a head or hand position already found nearby. If it does, the detected position is discarded and the latter is tried.

4 Experiment

Fig. 2 and Fig.3 show snapshots of e-Shadow and an example of animation frames used. For each input frame, detection results are interpreted and classified into predefined gestures, and then facial animations are injected into the head position.

Fig. 2. Snapshots of e-Shadow

Fig. 3. Animation frames in case of angry and joyful emotion

Table 1. presents experimental results using a casual webcam, where 60 different gestures were tested for each setting. Head positions were correctly detected in all the settings. *room1* is an office and *room2* is a houseroom. *s* is the scaling factor used to get an initial value of α from l. *type1* is the case that wrong hand positions were found

and *type2* is the case that hand positions were not found. When *s* is small, more *type1* errors occurred due to the cases that other body parts with small convex contours are mistaken as hand(s) and *type2* error occurred more when *s* is bigger. From the result, we conjecture that our system can work better in office environments than in house-rooms. Though the accuracies are relatively low when the distance between the user and camera is very near (0.8m), we can find that the system achieve acceptable performance by tuning *s*.

Table 1. Summary of the experimental result

distance	room	*s*=0.2			*s*=0.3			*s*=0.4		
		type1	type2	accuracy	type1	type2	accuracy	type1	type2	accuracy
0.8m	1	11	10	65.0	5	4	85.0	8	14	63.3
	2	16	7	61.7	13	4	71.7	6	18	60.0
1.5m	1	6	5	81.7	4	0	93.3	7	10	71.7
	2	14	4	70.0	7	5	80.0	4	11	75.0
3.0m	1	4	2	90.0	4	1	91.7	1	5	90.0
	2	6	4	83.3	5	3	86.7	3	4	88.3

5 Concluding Remarks

This paper presented e-Shadow, a real-time avatar, which can be utilized in web communication such as video chatting. The real shape of a user is obtained by a simplified background subtraction method and, a newly devised heuristic algorithm is adopted for 2D-based gesture detection. We demonstrated that the proposed avatar operates well under casual environments with some plausible settings of parameters.

Acknowledgement

This research was supported by Seoul Future Contents Convergence (SFCC) Cluster established by Seoul Industry-Academy-Research Cooperation Project.

References

1. Wang, L., Hu, W., Tan, T.: Recent Developments in Human Motion Analysis. Pattern Recognition, Vol. 36 (3) (2003) 585-601
2. Yang, Y.H., Levine, M.D.: The background primal sketch: an approach for tracking moving objects. Machine Vision and applications, Vol. 5 (1992) 17-34
3. Manohar, M., Ramapriyan, H.K.: Connected Component Labeling of Binary Image on a Mesh Connected Massively Parallel Processor. Vision, Computer Graphics and Image Processing, Vol. 45 (1989) 133-149
4. Wang, L., Hu, W., Tan, T.: Recent Developments in Human Motion Analysis. Pattern Recognition, Vol. 36 (3) (2003) 585-601

Two-Dimensional Discriminant Transform Based on Scatter Difference Criterion for Face Recognition

Cai-kou Chen[1,2] and Jing-yu Yang[2]

[1] Department of Computer Science and Engineering, Yangzhou University, 225001 Yangzhou, China
yzcck@126.com
[2] Department of Computer Science and Engineering, Nanjing University of Science and Technology, 210094 Nanjing, China
yangjy@public1.ptt.js.cn

Abstract. In this paper, a novel image discriminant analysis method, coined two-dimensional discriminant transform based on scatter difference criterion (2DSDD), is developed for image representation. The proposed 2DSDD scheme adopts the difference of both between-class scatter and within-class scatter as discriminant criterion. In this way, the small sample size problem usually occurred in the traditional Fisher discriminant analysis (LDA) is essentially avoided. In addition, the developed method directly depends on image matrices. That is to say, it is not necessary to convert the image matrix into high-dimensional image vector like those conventional linear discriminant methods prior to feature extraction so that much computational time would be saved. Finally, the experimental results on the ORL face database indicate that the proposed method outperforms Fisherfaces, the standard scatter difference discriminant analysis, not only in the computation efficiency, but also in its recognition performance.

1 Introduction

Recently, a novel LDA-like method called maximum scatter difference classifier (MSDC) [3] is proposed. Unlike LDA which bases on Fisher criterion [1,2], namely the ratio of between-class scatter and within-class scatter, MSDC adopts the difference of between-class scatter and within-class scatter as discriminant criterion for finding an optimal discriminant vector for classification. Compared with LDA methods, MSDC has two salient characteristics: (1) possessing the same physical meaning as LDA, (2) avoiding the singularity trouble of all LDA-like methods thoroughly. MSDC, however, extracts only an optimal projection vector for the resulting classification, which is insufficient for multi-class classification task. In the MSDC-based image recognition methods, the two-dimensional image matrices must be previously transformed into one-dimensional image vectors. The resulting image vectors usually lead to a high-dimensional image vector space, where it is difficult to evaluate the covariance matrix accurately due to its large size and the relatively small number of training samples. Recently, Yang et al. [4] proposed a new technique known as two-dimensional PCA, which directly bases on image matrix rather than image vector to calculate principal component.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 683 – 686, 2006.

In this paper, a novel image projection technique, coined two-dimensional discriminant transform based on scatter difference criterion (2DSDD), is developed for image feature extraction. As opposed to traditional MSDC and LDA, 2DSDD has two important advantages. First, it is based on 2D matrices rather than 1D vectors. As a result, less time is required to evaluate the covariance matrix accurately. Second, it is free to the inherent singularity problem for MSDC and LDA-like methods. Finally, 2DSDD has been used in face recognition. The experimental results are encouraging.

2 Basic Idea and Algorithm

Let $\mathbf{A}_i^j$ (i=1, 2, …, c, j=1, 2, …, N_i) be the jth $m \times n$ training image of the ith class, and c the number of all classes. We represent the number of the training samples per class by N_i. The mean images of the ith class and all training samples are denoted by $\bar{\mathbf{A}}_i$ and $\bar{\mathbf{A}}$, respectively.

Let $\mathbf{w}$ denote an n-dimensional unitary column vector, known as the projection axis. After projecting the training image $\mathbf{A}_i^j$ onto $\mathbf{w}$, the corresponding m-dimensional projected feature vector is derived as follows:

$$\mathbf{Y}_i^j = \mathbf{A}_i^j \mathbf{w}\,,\ i = 1,2,...,c\,,\ j = 1,2,...,N_i \tag{1}$$

To find the optimal projection direction $\mathbf{w}^*$ of the $\mathbf{A}_i^j$, we define the following discriminant criterion,

$$\boldsymbol{J}_{\mathbf{IS}}(\mathbf{w}) = tr(\mathbf{SB}) - tr(\mathbf{SW})\,, \tag{2}$$

where tr denotes the trace operator of a matrix, **SB** and **SW** denote the between-class scatter matrix and the within-class scatter matrix formed by the m-dimensional projected feature vectors, respectively,

$$\mathbf{SB} = \sum_{i=1}^{c} P(\omega_i)(\bar{\mathbf{Y}}_i - \bar{\mathbf{Y}})(\bar{\mathbf{Y}}_i - \bar{\mathbf{Y}})^T = \sum_{i=1}^{c} P(\omega_i)[(\bar{\mathbf{A}}_i - \bar{\mathbf{A}})\mathbf{w}][(\bar{\mathbf{A}}_i - \bar{\mathbf{A}})\mathbf{w}]^T\,, \tag{3}$$

$$\mathbf{SW} = \sum_{i=1}^{c} P(\omega_i)\frac{1}{N_i}\sum_{j=1}^{N_i}[(\mathbf{A}_i^j - \bar{\mathbf{A}}_i)\mathbf{w}][(\mathbf{A}_i^j - \bar{\mathbf{A}}_i)\mathbf{w}]^T\,. \tag{4}$$

From equation (3) and (4), it is easy to calculate the magnitude of between-class scatter and within-class scatter as follows,

$$tr(\mathbf{SB}) = \mathbf{w}^T\left(\sum_{i=1}^{c} P(\omega_i)(\bar{\mathbf{A}}_i - \bar{\mathbf{A}})^T(\bar{\mathbf{A}}_i - \bar{\mathbf{A}})\right)\mathbf{w}\,, \tag{5}$$

$$tr(\mathbf{SW}) = \mathbf{w}^T\left(\sum_{i=1}^{c} P(\omega_i)\frac{1}{N_i}\sum_{j=1}^{N_i}(\mathbf{A}_i^j - \bar{\mathbf{A}}_i)^T(\mathbf{A}_i^j - \bar{\mathbf{A}}_i)\right)\mathbf{w}\,. \tag{6}$$

We define the following two matrices,

$$\mathbf{IM}_b = \sum_{i=1}^{c} P(\omega_i)(\bar{\mathbf{A}}_i - \bar{\mathbf{A}})^T (\bar{\mathbf{A}}_i - \bar{\mathbf{A}}), \tag{7}$$

$$\mathbf{IM}_w = \sum_{i=1}^{c} P(\omega_i) \frac{1}{N_i} \sum_{j=1}^{N_i} (\mathbf{A}_i^j - \bar{\mathbf{A}}_i)^T (\mathbf{A}_i^j - \bar{\mathbf{A}}_i), \tag{8}$$

where $\mathbf{IM}_b$ and $\mathbf{IM}_w$ are referred to as *image between-class scatter matrix* and *image within-class scatter matrix*, respectively. By this definition, it is easy to show that they are non-negative definite matrices.

From equation (5), (6), (7), and (8), the equation (2) can be rewritten as

$$J_{IS}(\mathbf{w}) = \mathbf{w}^T \mathbf{IM}_b \mathbf{w} - \mathbf{w}^T \mathbf{IM}_w \mathbf{w} = \mathbf{w}^T (\mathbf{IM}_b - \mathbf{IM}_w)\mathbf{w}. \tag{9}$$

The new criterion function above, similar to the scatter difference criterion in equation (2), is called the *generalized scatter difference criterion*.

Let $\mathbf{w}_1, \cdots, \mathbf{w}_d$ denote a set of optimal projection axises obtained using 2DSDD. For a given training image $\mathbf{A}$, a set of the corresponding projected features,

$$\mathbf{Y}_k = \mathbf{A}\mathbf{w}_k, \ k = 1,2,...,d, \tag{10}$$

are obtained, which are called the *dominant discrimination components* (vectors) of $\mathbf{A}$. The dominant discrimination components are combined to construct a matrix, $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_d]$. The matrix $\mathbf{Y}$ is referred to as *discriminant matrix* or *discriminant image* of $\mathbf{A}$.

3 Experiments

The proposed 2DSDD method was used for face recognition and tested on the standard AT&T ORL face database. The ORL face database contains images from 40 individuals, each providing 10 different images. In our experiments, the whole database is divided into two parts evenly. One part is used for training and the other part is for test. In order to make full use of the available data and evaluate the generalization of the algorithm accurately, we adopt a cross-validation strategy and run the system 20 times. In each time, five images from each person are randomly selected as training samples, and the rest is for testing. Fisherfaces [1], MSDC, and the presented 2DSDD are employed for feature extraction. Finally, a nearest neighbor classifier is used for classification. The average CPU time consumed for training and testing and the top recognition rates of the above three methods are given in Table 1. Table1 shows 2DSDD achieves its maximal recognition rate of 96.5% and, it needs less CPU time compared to other methods. It should be noted that the speed difference between 2DSDD and other methods would become more significant with the increase of face database scale.

Table 1. The average CPU time (s) consumed for training and testing and the top recognition rates (%) of the four methods (CPU: PIII 800 MHZ, RAM: 256)

Method	Fisherfaces	MSDC	2DSDD
Recognition rate	89.5	89.5	96.5
CPU time	75.2	45.5	18.1

4 Conclusion

The paper has presented a new scheme for image feature extraction, namely scatter difference criterion-based two-dimensional discriminant transform (2DSDD), which is based on a straightforward projection of the two-dimensional image matrix according to scatter difference criterion. Since the size of the scatter matrices in 2DSDD is much smaller than those in the standard MSDC and Fisherfaces methods, less computation time is required for feature extraction. In the meantime, since the 2DSDD technique is based on scatter difference criterion, it is free to compute the inverse of the within-class matrix. Finally, experiments conducted on the ORL facial database have demonstrated that the 2DSDD features outperform those extracted by Fisherfaces, and MSDC methods in the performance of face recognition.

Acknowledgements

We wish to thank the National Science Foundation of China, under Grant No. 60472060, the University's Natural Science Research Program of Jiangsu Province under Grant No 05KJB520152, and the Jiangsu Planned Projects for Postdoctoral Research Funds for supporting this work.

References

1. P. N. Belhuneur, J. P. Hespanha, D. J. Kruebgman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Machine Intell. 19 (7) (1997) 711-720.
2. C. J. Liu, H. Wechsler, Robust coding schemes for indexing and retrieval from large face databases, IEEE Trans. Image Process. 9 (1) (2000) 132-137.
3. Fengxi Song, Shuhai Liu, Jingyu Yang, et al., Maximum scatter difference classifier and its application to text categorization, Computer Engineering. 31(5) (2005) 890-896.
4. Jian Yang, David Zhang, et al., Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Trans. Pattern Anal. Machine Intell. 26 (1) (2004) 131-137.

Hybrid Silhouette Extraction Method for Detecting and Tracking the Human Motion

Moon Hwan Kim[1], Jin Bae Park[1], In Ho Ra[2], and Young Hoon Joo[2]

[1] Yonsei University, Seodaemun-gu, 120-749, Seoul, Korea
{jmacs, jbpark}@control.yonsei.ac.kr
[2] Kunsan National University, Kunsan, 573-701, Chonbuk, , Korea
{irha, yhjoo)@kunsan.ac.kr

Abstract. Human motion analysis is an important research subject in human-robot interaction (HRI). However, before analyzing the human motion, silhouette of human body should be extracted from sequential images obtained by CCD camera. The intelligent robot system requires more robust silhouette extraction method because it has internal vibration and low resolution. In this paper, we discuss the hybrid silhouette extraction method for detecting and tracking the human motion. The proposed method is to combine and optimize the temporal and spatial gradient information. Also, we propose some compensation methods so as not to miss silhouette information due to poor images. Finally, we have shown the effectiveness and feasibility of the proposed method through some experiments.

1 Introduction

Human motion analysis is receiving an increasing attention for a wide spectrum of applications, such as man-machine interface, security surveillance, image retrieval and video indexing, and robot [1]. Especially, human motion analysis is studied as a new application of human-robot interaction (HRI) because it concerns with the key technique of HRI such as pose recognition and motion tracking of people.

There are two approaches to analyze human motion [1-11]; first, some bio-sensors fixed on the human joints send major bio-information to a main computer. After main computer analyzes the major positions and motion information of human body from the bio-information, it gives the motion paymasters of human body. The color marks can be used instead of bio-sensors. The main advantage of this approach is that it gives accurate features for motion analysis. However, in robot system or general situation, it is not adaptable because it requires the sensing equipment and the restricted environment. Second approach is to use the skeleton features of human body. To do this, the silhouette of human body should be first extracted from sequential images obtained by CCD camera. And human motion is then completed by analyzing the skeleton features based on silhouette of human body. It has been receiving an extensive attention from HRI researchers because it needs only one camera and does not need the main computer.

There are three conventional approaches to silhouette generation: temporal differencing [7], background subtraction [3-6], and optical flow [8]. The temporal

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 687 – 695, 2006.

differencing is very adaptive to dynamic environments, but it generally has a poor job for extracting all relevant feature pixels. The background subtraction provides accurate silhouette of human body, but it is extremely sensitive to change of images due to lighting and extraneous events. Optical flow can be used to detect features from sequential images with background motion, but most optical flow computation methods are very complex and are inapplicable to embedded system such as robot system. However, these conventional approaches are not adaptable to extract silhouette from images restricted within robot system.

To solve the above problems, we propose the hybrid silhouette extraction method for detecting and tracking the human motion. The proposed method is to fuse the temporal and spatial gradient method. Also, we propose some compensation methods so as not to miss silhouette information due to poor images. Finally, we show the effectiveness and feasibility of the proposed method through some experiments.

2 Hybrid Silhouette Extraction Method

Gradients represent instantaneous changes in the pixel value over time and space. The three types of gradients that have proven useful are depicted in each image pixel as followings; the spatial information in the horizontal direction and in the vertical direction, and the temporal information. At each image pixel, two gradient operators are applied to enhance both vertical differences and horizontal differences. A statistic is then calculated on a selected sub-region of the spatial gradient image to produce a scalar quantity. Thus, at each image pixel, we can obtain estimates of the magnitude and direction of the spatial gradient. Temporal gradients are calculated for each image pixel by subtracting pixel by pixel.

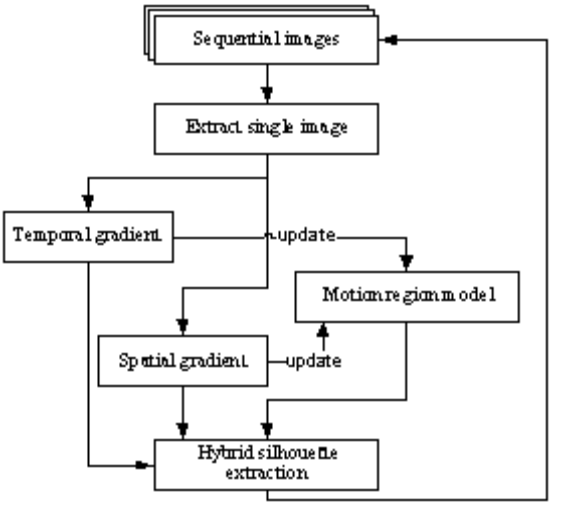


Fig 1. Hybrid silhouette extraction process

Using spatial gradient method, we can get all edge information of all things as well as a human body in one image. Using this edge information, it is not easy to find the unique silhouette of human body. On the other hand, using temporal gradient method, we can get the unique silhouette of human body from original image. But, this method has lost some information during updating the information obtained from the sequential images. To cope with these problems, we propose the hybrid silhouette extraction method. Figure 1 shows the block diagram for extracting the hybrid silhouette

2.1 Hybrid Silhouette Extraction

In this subsection, we discuss the hybrid silhouette extraction method. Let sequential image be defined as

$$\begin{aligned} I(x,y;t) &= [I_r(x,y;t), I_g(x,y;t), I_b(x,y;t)]^T \\ &= [p(t)]^T. \end{aligned} \tag{1}$$

Temporal gradient is temporal difference between consecutive two images. The temporal gradient and the spatial gradient including edge information are defined as

$$\begin{aligned} I_t &= \frac{\partial I}{\partial t} \\ I_s &= \left[\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}\right]. \end{aligned} \tag{2}$$

When we want to get the gradient information as key value of edge, it should be mapped into natural number. In this paper, we simply use the following mapping strategies:

$$\overline{I_t} = \|I_t\| \tag{3}$$

$$\overline{I_s} = \|I_s\| \tag{4}$$

In robot system, the temporal gradient $\overline{I_t}$ can have three kind of undesired values: 1) $\overline{I_t}$ can disappear because of camera problem. 2) $\overline{I_t}$ has insufficient silhouette information. 3) $\overline{I_t}$ has unnecessary motion information due to internal vibration of robot. To overcome first undesired situation, we should check sum of temporal gradient S_t defined as

$$S^t = \int_{\forall p \in I} f(\overline{I_t}, \gamma_t) dp \tag{5}$$

$$f(\overline{I_t}, \gamma_t) = \begin{cases} 1, & \text{If } \overline{I_t} > \gamma_t \\ 0, & \text{else} \end{cases}$$

where γ_t is minimum value to transform color image as binary image.

When S_t is zero, the old temporal gradient is replace to current temporal gradient.

The second and third undesired situation can be solved by using motion region model. To remove unnecessary motion information and complete insufficient silhouette information, we should use spatial gradient and history information of temporal gradient. History information of temporal gradient is the sum of consecutive some old temporal gradients. How to add spatial gradient and history information is key problem.

We propose the convex sum of temporal and spatial gradient can generate reliable accurate silhouette. The convex sum denotes as

$$\eta \overline{I_s}(x,y) + (1-\eta)\overline{I_t}(x,y) \tag{6}$$

where η is the convex sum parameter. By adjusting η, we can get accurate silhouette or imprecise silhouette. Generally, the convex sum parameter has static value. It can be determined by experimental or manual method, generally. We determine the convex parameter by using motion region model as shown in Figure 1.

The motion region model is defined as $R \subset \mathbb{R}^2$. The initial motion region model has zero. Then the motion region mode is updated as following, where γ_s is the binarization parameter of spatial image. $\overline{\gamma}_i$ and $\underline{\gamma}_i$ are upper and lower increasing parameters. $\overline{\gamma}_d$ and $\underline{\gamma}_d$ are upper and lower decreasing parameters. In subsection 2.2, the detailed description for parameters will be presented. Finally, the hybrid silhouette is calculated as

$$\begin{aligned} R(x,y;t+1) &= \begin{cases} R(x,y;t)+\gamma_i, \overline{I_t}(x,y;t+1) > \gamma_t \\ R(x,y;t)+\gamma_d, \overline{I_t}(x,y;t+1) < \gamma_t \end{cases} \\ \gamma_i &= \begin{cases} \overline{\gamma}_i, \overline{I_s}(x,y;t+1) > \gamma_s \\ \underline{\gamma}_i, \overline{I_s}(x,y;t+1) < \gamma_s \end{cases} \\ \gamma_d &= \begin{cases} \overline{\gamma}_d, \overline{I_s}(x,y;t+1) > \gamma_s \\ \underline{\gamma}_d, \overline{I_s}(x,y;t+1) < \gamma_s \end{cases} \end{aligned} \tag{7}$$

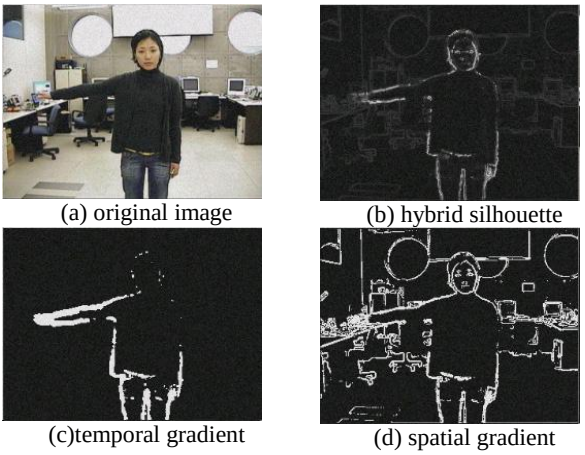
$$\begin{aligned} \overline{I_b}(x,y) &= \eta \overline{I_s}(x,y) + (1-\eta)\overline{I_t}(x,y) \\ \eta &= \begin{cases} \eta_R, R(x,y) > \gamma_R \\ \eta_B, R(,xy) < \gamma_R \end{cases} \end{aligned} \tag{8}$$

where η_R is convex sum parameter for motion region. η_B is the convex sum parameter for background. Determination of these parameters is also discussed in section 2.2.

2.2 Choices of Parameters

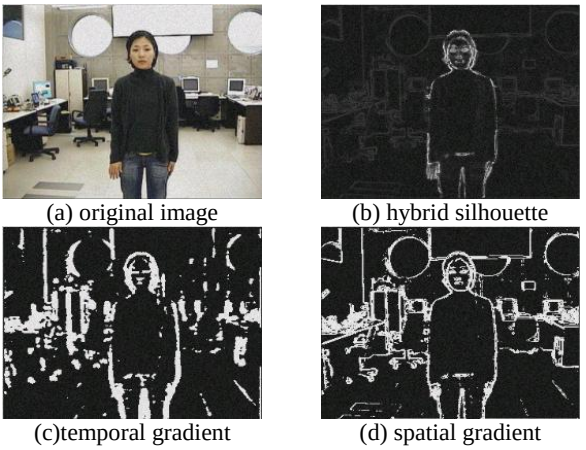
There exist many parameters in hybrid silhouette extraction method proposed in this paper. Unfortunately, these parameters are not easy to be determined via automatic or intelligent method. Therefore, we provide a guideline on appropriate relative choices of parameters. The upper and lower increasing parameters have the following constraint

$$\overline{\gamma}_i > \underline{\gamma}_i > 0 \tag{9}$$



(a) original image (b) hybrid silhouette (c)temporal gradient (d) spatial gradient

Fig. 2. Hybrid silhouette extraction: $\overline{I}_t$ has insufficient silhouette



(a) original image (b) hybrid silhouette (c)temporal gradient (d) spatial gradient

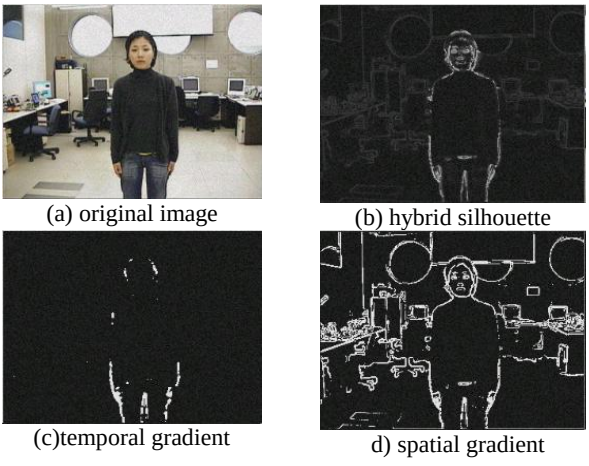
Fig. 3. Hybrid silhouette extraction: $\overline{I}_t$ has unnecessary motion information

Similarly, the upper and lower decreasing parameters have the following constraint

$$0 < \overline{\gamma}_d < \underline{\gamma}_d \tag{10}$$

The increasing and decreasing parameters have to comply to the following constraint

$$\begin{aligned} \left|\overline{\gamma}_i\right| &> \alpha \left|\overline{\gamma}_d\right| \\ \left|\underline{\gamma}_i\right| &< \left|\underline{\gamma}_d\right| \end{aligned} \tag{11}$$



(a) original image (b) hybrid silhouette

(c)temporal gradient d) spatial gradient

Fig. 4. Hybrid silhouette extraction: $\overline{I}_t$ is disappeared

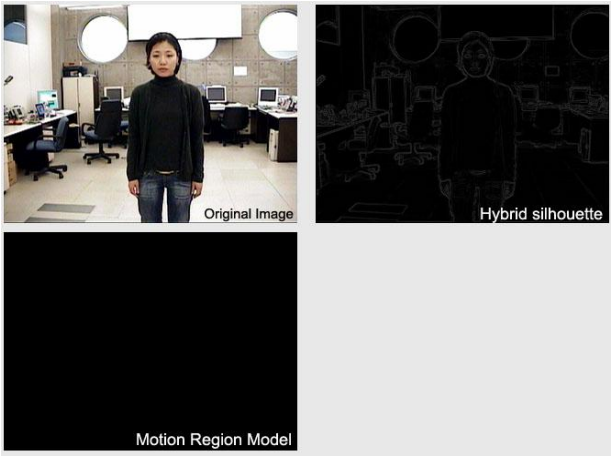


Fig. 5. Changes of hybrid silhouette of frame 1 and motion region model

where α means temporal memory length. When α is two, the motion region model have recent two past motion information. The convex sum parameter η_R and η_B have following constraint

$$\eta_R > 0.5 > \eta_B \tag{12}$$

Large $\overline{\gamma}_i$ values lead to make spatial gradient complete insufficient silhouette whereas small $\overline{\gamma}_i$ values yield less complete silhouette. However, large $\overline{\gamma}_i$ values also yield redundant silhouette information.

3 Experimental Results

We demonstrate the feasibility of the proposed method through some experiments. The sequential images are obtained by CCD color camera fixed on the robot. The sequential images are captured as 320x240 resolutions with 24 bit color depth. The proposed silhouette extraction method is applied to sequential images. The parameters used in this experiment were manually chosen. We performed some experiments for thee undesired situations as described in Section 2.1. Figures 2-4 show the results of the proposed extraction method under three undesired situations. Each figure has 4 pictures containing original image, temporal gradient, spatial gradient, and hybrid silhouette using one image. As shown in these figures, we could confirm the accurate hybrid silhouette in spite of unnecessary or insufficient silhouette information.

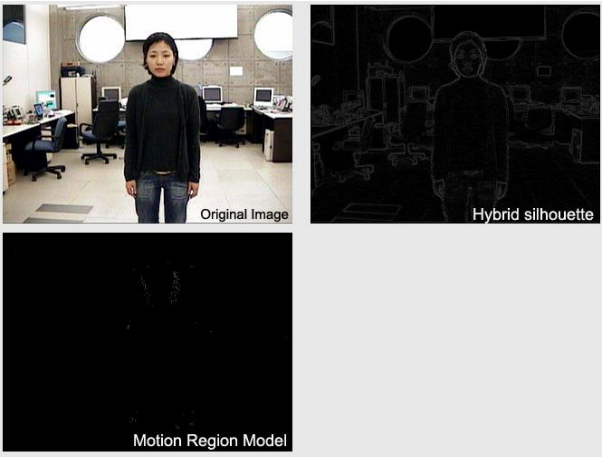


Fig. 6. Changes of hybrid silhouette of frame 2 and motion region model

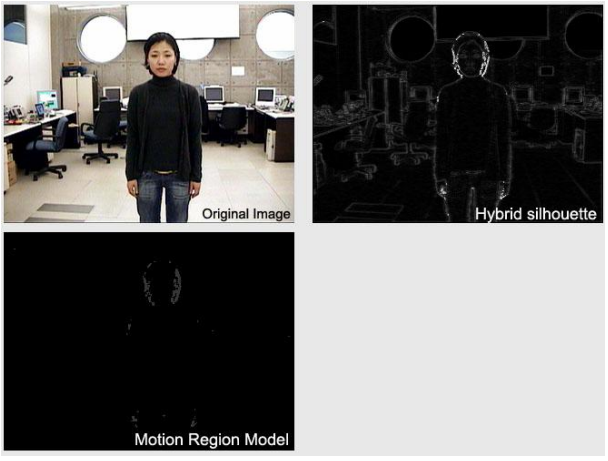


Fig 7. Changes of hybrid silhouette of frame 3 and motion region model

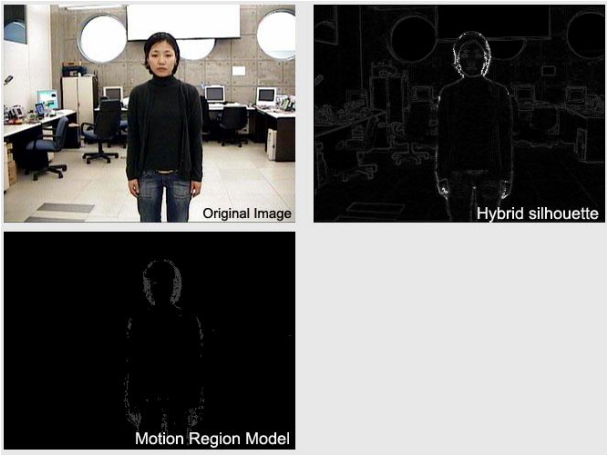


Fig 8. Changes of hybrid silhouette of frame 4 and motion region model

Figure 5-8 shows seven updating results of motion region model and hybrid silhouette obtained from seven sequential images. Each figure shows 3 pictures containing original image, image obtained by motion region model, and hybrid silhouette extracted by the proposed method. As shown in this Figure 5-8, we can confirm more accurate hybrid silhouette after updating the information obtained from the sequential images.

4 Conclusion

In this paper, we proposed the hybrid silhouette extraction method of human body for detecting and tracking the human motion. The proposed method is to combine and optimize the temporal and spatial gradient information. Also, we proposed some

compensation methods so as not to miss silhouette information due to poor images. Finally, we have shown the effectiveness and feasibility of the proposed method through some experiments.

Acknowledgment

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD (KRF-2005-042-D00268).

References

1. Aggarwal, J. K., Cai, Q.: Human Motion Analysis: A Review. Computer Vision and Image Understanding 73 (1999) 428—440
2. Fan, B., Wang, Z.-F.: Pose estimation of human body based on silhouette images. International Conference on Information Acquisition Proceedings, (2004) 296-300
3. Haritaoglu, I., Cutler, R., Hawood, D., Davis, L.: Backpack: Detection of people carrying objects using silhouettes. Computer Vision and Image Understanding, 3 (2001) 385-397
4. Haritaoglu, I., Harwood, D., Davis, L.: A real time system for detection and tracking people. Journal of Image and Vision Computing, 7 (1999) 345-352
5. Haritaoglu, I., Hawood, D., Davis, L.: Who? When? Where? What? A Real Time System for Detecting and Tracking People, Automatic Face and Gesture Recognition, 3 (1998) 222-227
6. Blake, A., Isard, M., Reynard, D.: Learning to track curves in motion of contours. Artificial Intelligence, 78 (1995) 101-133
7. Anderson, C., Burt, P., van der Wal, G.: Change detection and tracking using pyramid transformation techniques. In Proceedings of SPIE - Intelligent Robots and Computer Vision, 579 (1985) 72-78
8. Barron, J., Fleet, D., Beauchemin, S.: Performance of optical flow techniques, Int. Journal of Computer Vision. 12 (1994) 42-77
9. Bobick, A., Davis, J., Intille, S., Baird, F., Cambell, L., Irinov, Y., Pinhanez, C., Wilson, A.: KidsRoom: Action Recognition In An Interactive Story environment. M.I.T.TR. 398 (1996)
10. Fejes, S., Davis, L.S.: Exploring Visual Motion Using Projections of Flow Fields. In. Proc. of the DARPA Image Understanding Workshop, New Orleans,LA (1997) 113-122
11. Ju, S., Black, M., Yacoob, Y.: Cardboard People: A Parameterized Model of Articulated Image Motion. Int. Conference on Face and Gesture Analysis, (1996) 38-44.

Two-Dimensional PCA Combined with PCA for Neural Network Based Image Registration

Anbang Xu, Xin Jin, and Ping Guo*

Image Processing & Pattern Recognition Laboratory, Beijing Normal University
Beijing 100875, China
anbangxu@mail.bnu.edu.cn, xinjin796@126.com, pguo@ieee.org

Abstract. A novel image registration scheme is proposed. In the proposed scheme, two-dimensional principal component analysis (2DPCA) combined with principal component analysis (PCA) is used to extract features from the image sets and these features are fed into feedforward neural networks to provide translation, rotation and scaling parameters. Comparison experiments between 2DPCA combined with PCA based method and the other two former methods: discrete cosine transform (DCT) and Zernike moment, are performed. The results indicate that the proposed scheme is both accurate and remarkably robust to noise.

1 Introduction

Image registration is the process of aligning two or more images of the same scene. Image registration techniques are embedded in a lot of visual intelligent systems, such as robotics, target recognition, remote medical treatment and autonomous navigation. The common image registration methods are divided into two types: intensity-based methods and feature-based methods. The analysis and evaluation for various techniques and methods of image registration are carried out on the basis of these two sorts, while the feature-based methods are emphasized [1].

Recently, Itamar Ethanany [2] proposed to use feedforward neural network (FNN) to register a distorted image through 144 Discrete Cosine Transform (DCT) -base band coefficients as the feature vector. But this method has too large lumber of input feature vectors for the un-orthogonality of DCT based space, thus expose low computational efficiency and high requirements on computer performance. Later, Wu and Xie [3] used low order Zernike moments instead of DCT coefficients to register affine transform parameters. Although the input vector dimension of neural networks is reduced, the estimation accuracy is still not satisfied.

A key step in these image registration schemes is to extract image features to form the FNN training data set. The main challenge in this step is how to reduce the high dimensional input data and retain the feature for image registration. In this paper, we develop a novel method to image registration, which uses 2DPCA combined with PCA for feature extraction and then feed these feature vectors into a feedforward neural network to obtain register affine transform parameters. Experimental results

* Corresponding author.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 696 – 705, 2006.

show that the scheme we proposed is better than other methods in terms of accuracy and robustness to noise.

This paper is organized as follows: In section 2, the new image registration scheme and its algorithm are presented. Section 3 focuses on experimental results comparing with other methods under various noisy conditions. Finally, the conclusions are presented in section 4.

2 2DPCA Combined with PCA and FNN Based IMAGE Registration Scheme

The registration scheme consists of two stages: the pre-registration phase and the registration phase. In the pre-registration phase, first, a training set is synthesized by the reference image. The feature coefficients are extracted from the training set with the method of 2DPCA combined with PCA, and then these feature coefficients are fed as inputs to a feedforward neural network. Second, the neural network is trained and its target outputs are affine parameters. In the registration phase, as the neural network is trained, the remainder work is simple. We just use the same method to extract features from the registered image and feed these features to the trained network to get the affine parameters estimated. The registration approach is shown in Fig. 1 in detail.

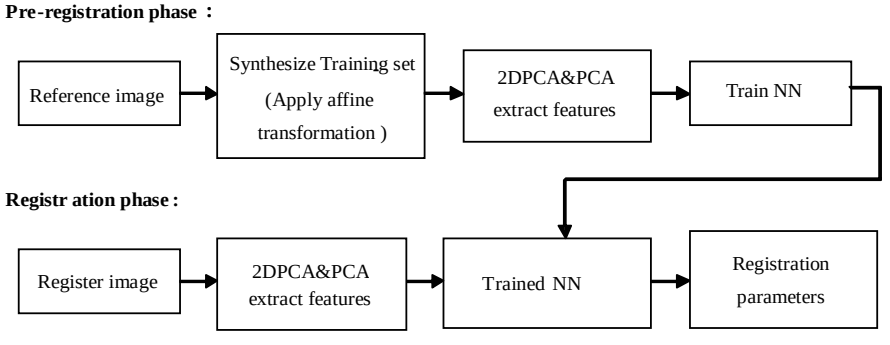


Fig. 1. The pre-registration and registration phases of the proposed scheme

2.1 Affine Transformation

Geometrical transformation can be represented in many different ways, affine transformation is one of the most common used transformations. It is a kind of rigid transformation, having many good mathematics characteristics. We can represent images with 2D matrices. Affine transformation can be described by the following matrix equation [2], [3]:

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} t_x \\ t_y \end{pmatrix} + s \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}. \tag{1}$$

In the equation, there are four basic parameters for the transformation, where (x_1,y_1) denotes the original image coordinate, (x_2,y_2) denotes the transformed image coordinate in another image, t_x, t_y are the translation parameters, s is a scaling factor and θ is a rotation angle. In this paper, we will adopt this transformation model.

2.2 Feature Extraction Methods

2.2.1 Principal Component Analysis

PCA is a classical multivariate data analysis method that is useful in linear feature extraction and data compression [4]. It is essentially equivalent to Karhunen-Loeve transformation and closely related to factor analysis. All these methods are based on 2nd-order statistics of the data.

The PCA finds a linear transformation $\mathbf{y}=\mathbf{W}\mathbf{x}$ such that the retained variance is maximized. It can be also viewed as a linear transformation that minimizes the reconstruction error [5]. Each row vector of $\mathbf{W}$ corresponds to the normalized orthogonal eigenvector of the data covariance matrix.

One simple approach to PCA is to use singular value decomposition (SVD). Let us denote the data covariance matrix by $\mathbf{R}_\mathbf{x}(0) = E\{\mathbf{x}(t)\mathbf{x}^T(t)\}$. Then the SVD of $\mathbf{R}_\mathbf{x}(0)$ gives $\mathbf{R}_\mathbf{x}(0) = \mathbf{U}\mathbf{D}\mathbf{U}^T$, where $\mathbf{U}= [\mathbf{U}_s,\mathbf{U}_n]$ is the eigenvector matrix (i.e. modal matrix) and $\mathbf{D}$ is the diagonal matrix whose diagonal elements correspond to the eigenvalues of $\mathbf{R}_\mathbf{x}(0)$ (in descending order). Then the PCA transformation from *m*-dimensional data to *n*-dimensional subspace is given by choosing the first *n* column vectors, i.e., *n* principal component vector $\mathbf{y}$ is given by $\mathbf{y}=\mathbf{U}_s^T\mathbf{x}$.

2.2.2 Two-Dimensional PCA

Two-dimensional principal component analysis [6] is based on 2D matrices rather than 1D vectors. It is mean that the image matrix does not need to be previously transformed into a vector. An image covariance matrix can be computed directly by using the original image matrices. So, 2DPCA has two advantages over PCA. First, it is easier to compute the covariance matrix accurately. Second, less time is needed to determine the corresponding eigenvectors.

We define $\mathbf{X}$ as an n-dimensional unitary column vector and present image $\mathbf{A}$ as an $m \times n$ random matrix. The main idea is to project the $m \times n$ random matrix, onto $\mathbf{X}$ by the following linear transformation [6], [7], [8]:

$$\mathbf{Y} = \mathbf{A}\mathbf{X}, \tag{2}$$

Where $\mathbf{Y}$ is an m-dimensional projected vector which is called the projected feature vector of image $\mathbf{A}$. In order to determine a proper projection vector $\mathbf{X}$, we introduce the total scatter of the projected samples to measure the discriminatory power of the projection vector $\mathbf{X}$. The total scatter of the projected samples can be computed as the following equation:

$$J(\mathbf{X}) = \mathrm{tr}(\mathbf{S}_x), \tag{3}$$

where $\mathbf{S}_x$ is the covariance matrix of the projected feature vectors of the training samples and tr($\mathbf{S}_x$) denotes the trace of $\mathbf{S}_x$. The covariance matrix $\mathbf{S}_x$ can be computed by

$$\begin{aligned}\mathbf{S}_x &= E(\mathbf{Y}-E\mathbf{Y})(\mathbf{Y}-E\mathbf{Y})^T = E[\mathbf{AX}-E(\mathbf{AX})][\mathbf{AX}-E(\mathbf{AX})]^T \\ &= E[(\mathbf{A}-E\mathbf{A})\mathbf{X}][(\mathbf{A}-E\mathbf{A})\mathbf{X}]^T .\end{aligned}$$

So,

$$\operatorname{tr}(\mathbf{S}_x) = \mathbf{X}^T \left[E(\mathbf{A}-E\mathbf{A})^T (\mathbf{A}-E\mathbf{A}) \right] \mathbf{X}. \tag{4}$$

Defining the following matrix **H** which is called the image covariance (scatter) matrix:

$$\mathbf{H} = E\left[(\mathbf{A}-E\mathbf{A})^T (\mathbf{A}-E\mathbf{A}) \right], \tag{5}$$

where **H** is an n × n nonnegative definite matrix from its definition. The matrix **H** can be computed directly by using the training samples. Given training image samples $\mathbf{A}_1,\ldots,\mathbf{A}_M$, the jth training image can be denoted by an $m \times n$ matrix $\mathbf{A}_j$ (j =1, 2,…, M), and the average image of all training samples is $\overline{\mathbf{A}}$. Then, **H** can be computed by the following equation:

$$\mathbf{H} = \frac{1}{M} \sum_{j=1}^{M} (\mathbf{A}_j - \overline{\mathbf{A}})^T (\mathbf{A}_j - \overline{\mathbf{A}}). \tag{6}$$

The criterion in (3) can be represented by

$$J(\mathbf{X}) = \mathbf{X}^T \mathbf{H} \mathbf{X}, \tag{7}$$

where **X** is a unitary column vector. We can call this criterion as the generalized total scatter criterion. The unitary vector **X** that maximizes the criterion is called the optimal projection axis. Intuitively, this means that the total scatter of the projected samples is maximized after the projection of an image matrix onto **X**.

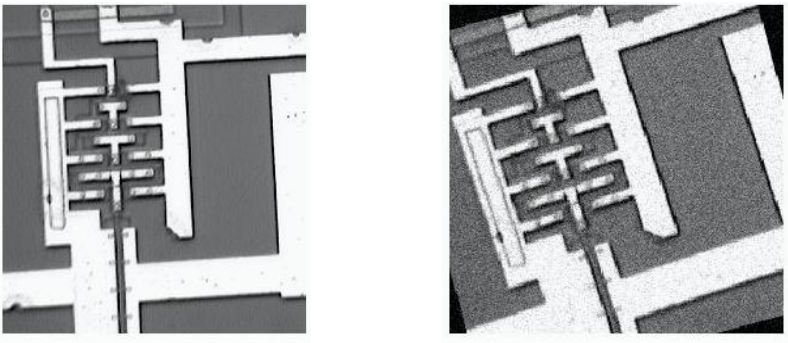
In fact, the optimal projection axes, $\mathbf{X}_1, \mathbf{X}_2,\ldots,\mathbf{X}_d$, are the orthonormal eigenvectors of **H** corresponding to the first d largest eigenvalues. So we can adopt these projection vectors, $\mathbf{X}_1,\ldots,\mathbf{X}_d$, for feature extraction. For a given image sample **A**, let

$$\mathbf{Y}_k = \mathbf{A}\mathbf{X}_k, k = 1, 2, ..., d . \tag{8}$$

Where $\mathbf{Y}_k$ is the principal component (vectors) of the sample image **A**. So, a series of projected feature vectors, $\mathbf{Y}_1,\ldots,\mathbf{Y}_d$, can be obtained. From the above equation, we can see that each principal component of 2DPCA is a vector, comparing with the principal component of PCA which is a scalar.

As the principal component vectors are obtained from the above equation, an $m \times d$ matrix **B** can be defined as [$\mathbf{Y}_1,\ldots, \mathbf{Y}_d$]. The matrix B is the feature matrix of the image sample **A**.

However, it should be pointed out that 2DPCA-based image representation was not as efficient as PCA in terms of storage requirements, since 2DPCA requires more coefficients for image representation than PCA [6]. So in the proposed image registration scheme, we take advantage of PCA after 2DPCA for further dimensional reduction.



(a) (b)

Fig. 2. (a) An original image and (b) a registered image in the training set with 15-degree rotation, 110% scaling, translation of 2 pixel and -3 pixel on X-axis and Y-axis respectively at a signal-to-noise ratio (SNR) of 20 dB

2.3 Image Registration Scheme with FNN

The image registration scheme includes training the FNN to provide the required affine parameters. Each image in the training set is generated by applying an affine transformation. The affine parameters are randomly changed in a predefined range so as to reduce correlations among images. In order to improve the generalization and immunity of the FNN from over-sensitivity to distorted inputs, we introduce noise in the image synthesis. Then we employ 2DPCA combined with PCA as a feature extraction mechanism presented to the FNN.

Because there is no theoretical guidance to decide the number of neurons in hidden layer [9] [10] and the low computational efficiency of some methods such as cross

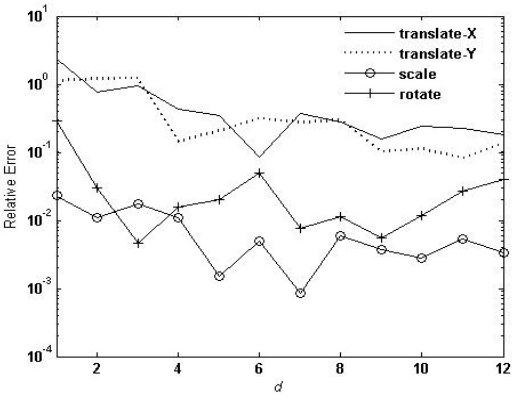


Fig. 3. Relative errors with respect to translation, scale and rotation vary with the number of 2DPCA component feature vectors (*d*)

validation and bootstrap, for the FNN with three layer structure we can use the empirical formula [11] to compute the hidden neuron number:

$$p = \sqrt{0.43mn + 0.12n^2 + 2.54m + 0.77n + 0.35} + 0.5, \quad (9)$$

where p is the number of neurons in hidden layer, m is the number of input neurons and n is the number of output neurons. Here the structure of the FNN is that contains 60 inputs, 4 outputs and 17 hidden neurons. Sigmoid transfer functions are employed in the hidden layers while linear functions characterize the output-level neurons. The FNN is trained by using the Levenberg-Marquardt algorithm [12] based on gradient-descent learning technique.

3 Experiments

3.1 Parameter Selection and Accuracy of the Proposed Method

In the experiment, a pair of 256×256 resolution images was used. Fig. 2 shows one of the original images and a transformed image containing translation, rotation and scaling.

The training set consists of 300 images, each image is transformed from the reference image by translating, rotating and scaling randomly within a predefined range. Besides, additive Gaussian noise and Salt & Pepper type noise are applied on each image in various intensities. We can also generate some test samples to demonstrate the registration accuracy of the proposed method. The optimal projection vectors of 2DPCA, $\mathbf{X}_1,\ldots,\mathbf{X}_d$, described in (8) are used for feature extraction, and then we get the $256 \times d$ feature matrix from each image in the given training set. After using 2DPCA to feature extraction, we use PCA for further dimensional

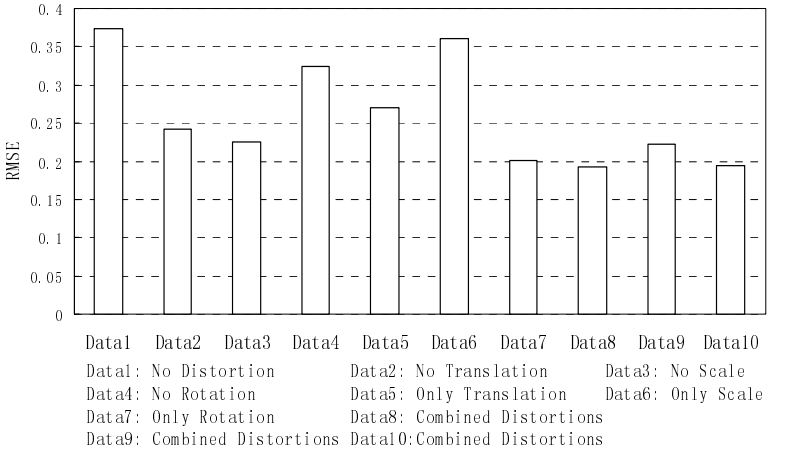


Fig. 4. RMSE between the registered image and the original image under 15 db SNR

Table 1. Distortion parameters estimation Under 20 db SNR

Distortions Parameters			Registered Parameters
No Distortion	Translation	[0,0]	[-0.39192, -0.2638]
	Scale	1	1.0155
	Rotation	0	0.08075
No Translation	Translation	[0,0]	[0.27772, 0.5015]
	Scale	0.9	0.90003
	Rotation	21	20.95
No Scale	Translation	[-2,1]	[-2.4469, 0.68821]
	Scale	1	0.99839
	Rotation	3	2.7827
No Rotation	Translation	[5,-5]	[4.7748, -4.4004]
	Scale	0.75	0.75554
	Rotation	0	-0.14539
Only Translation	Translation	[-5,5]	[-4.6695, 4.438]
	Scale	1	0.99888
	Rotation	0	-0.14539
Only Scale	Translation	[0,0]	[0.46742, 0.44583]
	Scale	1.25	1.2357
	Rotation	0	0.64803
Only Rotation	Translation	[0,0]	[-0.3987, -0.2838]
	Scale	1	1.0022
	Rotation	38	38.609
Combined Distortions	Translation	[-5,4]	[-5.1254, 3.8283]
	Scale	0.9	0.92162
	Rotation	25	24.919
Combined Distortions	Translation	[5, -1]	[4.7844, -1.265]
	Scale	1.134	1.1324
	Rotation	11.2	11.218
Combined Distortions	Translation	[-2, -3]	[-2.2719, -3.283]
	Scale	0.7	0.71089
	Rotation	33	33.011

reduction and the final the number of input feature is 60. These feature coefficients of images are inputs of FNN, the FNN is trained and its outputs are affine parameters. Finally, the feature coefficients are extracted from the registered image with the same method and fed as inputs into the trained neural network to get the affine parameters estimated.

In this experiment, as we add the d from 1 to 12, Fig. 3 shows the relative error varies with d. As observed in Fig. 3, the relative error is reduced generally with the increase of d. we can also see that when $d > 4$, the registration scheme we proposed becomes accurate in estimating the affine parameters especially in estimating the scale parameter and the rotate parameter. The accuracy became steady and satisfying as $d \geq 10$.

Simulation results are shown in Table 1 as d = 12 and SNR = 20 db. The accuracy of parameter estimation can be also evaluated by root mean square error (RMSE) between the registered image and the original image, Fig. 4 shows the results under various test data sets under 15 db SNR.

3.2 Comparison with Other Methods Under Different Noisy Conditions

In the proposed method, we choose d=10 and dimension of input feature is 60.While in Zernike moment-based method, the dimension of input feature is 60 and 17 neurons in hidden layer, and in DCT-based method, the dimension of input feature is =64 and 18 neurons in hidden layer. Same with the above experiment, a training set consisting of 300 images is synthesized. In order to evaluate registration performance with Gaussian noise, we take 40 images for each the evaluated SNR value. The test image as shown in Fig. 1 (a) is rotated 15 degree, 110% scaled, translated 2 pixels and -3 pixels on X-axis and Y-axis respectively. Fig. 5 depicts the results of estimating the affine transform parameters under different SNR values.

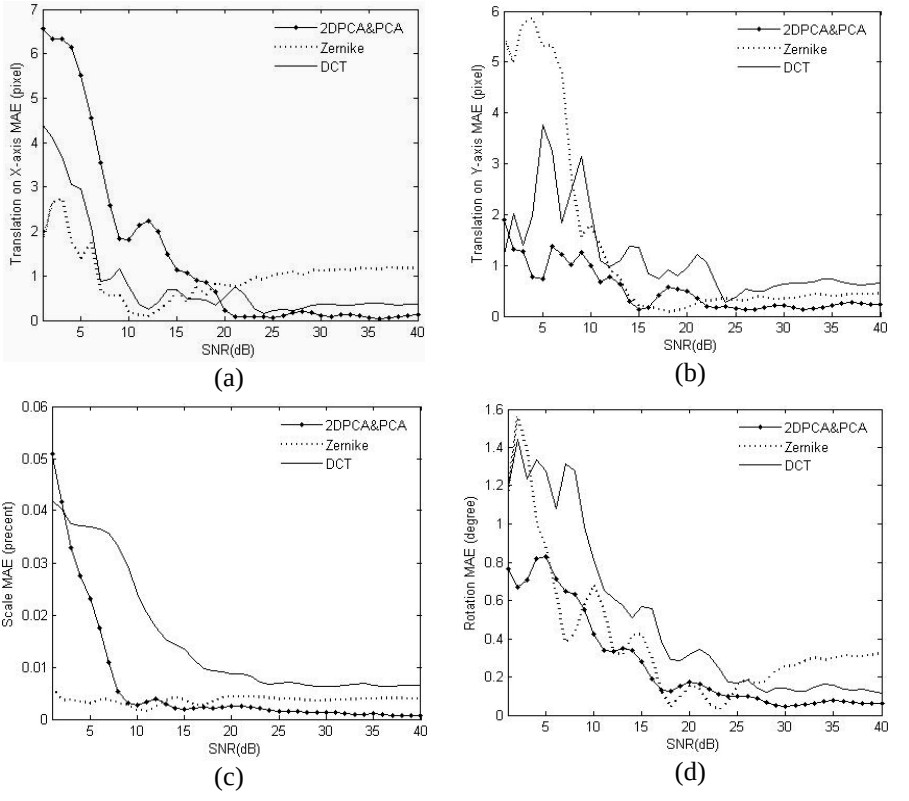


Fig. 5. Absolute registration error comparison DCT, Zernike and 2DPCA combined with PCA. (a) error with respect to translation on X-axis, (b) translation on Y-axis, (c) scale and (d) rotation errors.

By analyzing the experiment results, it can be found that our method shows more accurate than the other two methods, especially when SNR is larger than 20 dB.

4 Conclusions

In this paper, a novel image registration scheme is proposed, which adopts the FNN and 2DPCA combined with PCA to register affine transform parameters. Comparison experiments for feature extraction based image registration among DCT, Zernike moment and 2DPCA combined with PCA are performed under different noisy conditions. It is shown that the proposed scheme has more accurate registration performance and robust to noise than the other methods. The experiment results suggest that 2DPCA combined with PCA is a much efficient method to represent image registration feature.

The proposed method still deserves further study. First, the performance of this method is not very satisfying when SNR is low, the solution may be that other feature extraction methods [13], [14] are investigated. Second, the generalization abilities of the FNN can be optimized in other methods such as regularization [15]. Third, 2D-image registration problem is not satisfied for some practical applications. In the further study, we will engage to find the solution for these problems and try to develop 3D-image registration based on the proposed method.

Acknowledgement

The research work described in this paper was fully supported by a grant from the National Natural Science Foundation of China (Project No. 60275002).

The authors would like to thank anonymous reviewers for their useful comments and suggestions.

References

1. Tan, Y.P., Yap, K.H., Wang, L.P. (Eds.): Intelligent Multimedia Processing with Soft Computing, Springer-Verlag Heidelberg, 2004
2. Elhanany, I., Sheinfeld, M., Beckl, A., Kadmon, Y., Tal, N., Tirosh, D.: Robust image registration based on feedforward neural networks. In: IEEE International Conference on System, Man and Cybernetics, Vol.2 (2000) 1507–1511
3. Wu, J., Xie, J.: Image Registration Scheme Utilizing Feedforward Neural Networks. In: Proceedings of the 5th World Congress on Intelligent Control and Automation, Vol. 5 (2004) 4046–4048
4. Zeng, X.Y., Chen, Y.W., Nakao, Z., Lu, H.: A new texture feature based on PCA maps and its application to image retrieval. IEICE Trans. Inf. and Syst. **86**(5) (2003) 929–936
5. Diamantaras, K.I. and Kung, S.Y.: Principal Component Neural Networks: Theory and Applications. John Wiley & Sons, INC Vol. 1 (1996)
6. Yang, J., Zhang, D., *et al.*: Two-dimensional PCA a new approach to appearance-based face representation and recognition. IEEE Trans. Pattern Analysis and Machine Intelligence **26**(6) (2004) 131–137

7. Liu, K., *et al*.,:Algebraic Feature Extraction for Image Recognition Based on an Optimal Discriminant Criterion. Pattern Recognition **26**(2) (1993) 903–911
8. Yang, J., Yang, J.Y.: From Image Vector to Matrix: A Straightforward Image Projection Technique—IMPCA vs. PCA. Pattern Recognition **35**(7) (2002) 1997–1999
9. Ravdin P.M., Clark, G.M., Hilsenbeck, S.G., *et al*: A demonstration that breast cancer recurrence can be predicted by neural network analysis. Breast Cancer Research and Treatment **21**(1) (1992) 47–53
10. Laurentiis, D.M., Ravdin, P.M.: Survival analysis of censored data: Neural network analysis detection of complex interaction between variables. Breast Cancer Research and Treatment **32**(1) (1994) 113–118
11. White, H.: Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mapping. Neural Networks **3**(5) (1990) 535–549
12. Hagan, M.T., Menhaj, M.: Training Feedforward Networks with Marquardt Algorithm. IEEE Trans. Neural Networks **1**(1) (1994) 113–118
13. Xu, A.B., Guo, P.: Isomap and Neural Networks based Image Registration Scheme. In: International Symposium on Neural Networks (ISNN 2006), Lecture Notes in Computer Science, Vol. 3972. Springer- Verlag, Berlin Heidelberg (2006) 486–491
14. Xu, A.B., Jin, X., Guo, P, Bie, R.F.: KICA Feature Extraction in Application to FNN based Image Registration. Will appear in: Proceedings of the 2006 International Joint Conference on Neural Networks, (2006)
15. Guo, P., Lyu, M. R., Philip Chen, C. L.: Regularization Parameter Estimation for Feedforward Neural Networks. IEEE Trans. Neural Networks **33**(1) (2003) 35–44

SAR Speckle Reduction Based on Undecimated Tree-Structured Wavelet Transform

Ying Li, Jianglin Yang, Li Sun, and Yanning Zhang

School of Computer Science, Northwest Polytechnical University, Xi'an, 710072, China
lybyp@nwpu.edu.cn

Abstract. This paper proposes a novel filtering method for removing such speckle noise from Synthetic Aperture Radar image that combines the Stationary Tree-structured Wavelet Transform (STWT) with a Bayesian wavelet estimator. Experimental results on several test images by using the proposed method show that, the proposed method yields significantly superior image quality and better Peak Signal to Noise Ratio (PSNR).

1 Introduction

Speckle reduction is an important issue in SAR image processing. In recent year, the wavelet-based techniques have been widely utilized in SAR image processing [1–4]. Wavelet denoising methods usually apply *hard* or *soft* thresholding to the detailed wavelet coefficients of noisy images [5]. However, most of the wavelet thresholding methods suffer from the drawback that chosen threshold may not match the specific distribution of signal and noise components in different scales. To address this issue, the non-linear estimators based on formal Bayesian theory are developed [3], [4].

In this paper, we propose an efficient Stationary Tree-structured Wavelet Transform (STWT) based despeckling method that preserves edges of the original image. We avoid the log-transform and derive a novel wavelet shrinkage factor based on the special form of the generalized Gaussian distribution, i.e. Laplace distribution model, of wavelet coefficients.

2 Stationary Tree-Structure Wavelet Transform

Generally there are two types of wavelet decomposition namely Discrete Wavelet Transform (DWT) and Stationary Wavelet Transform (SWT). Both DWT and SWT are the traditional pyramid-type wavelet transforms which recursively decompose subimages into low frequency channels. However, since the most significant information of texture often appears in the middle frequency channels, recursive decomposition just into the lower frequency region may not be optimum for SAR images, which contain a great deal texture information.

Here we approach this problem by analyzing SAR images by the stationary tree-structured wavelet transform (STWT). The key point in our approach is that we no longer simply apply the decomposition to the low frequency subimages only. Instead,

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 706 – 709, 2006.

decomposition can be applied to the output of any filter h_{LL}, h_{LH}, h_{HL}, or h_{HH}, which is found to contain significant information. At the same time, the down sampling process is not applied to make it shift invariant. Note that it is usually unnecessary and expensive to decompose all subimages in each scale to achieve a full decomposition. To avoid a full decomposition, we may consider a criterion, such as l_1 norm or norm, to decide whether a decomposition is needed for a particular output. The STWT is given below.

1) Decompose an image using SWT into four subimages.
2) Calculate the average energy of each decomposed subimage.

$$e = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\left|x(m,n)\right|^2 \tag{1}$$

3) If the energy of a subimage is significantly smaller than others of the same level, we stop the decomposition in this region since it contains less information.
4) If the energy of a subimage is significantly larger, we apply the above decomposition procedure to the subimage once again.

3 Bayesian Wavelet Estimate

In the wavelet domain, multiplicative speckle noise can be converted into additive noise without the use of the log-transform [6]-[8]

$$W_X = W[X] = W[YF] = W[Y] + W[Y(F-1)] = W_Y + W_N \tag{2}$$

Equation (2) can be simply formulated as $y = w + n$, where *y* noisy wavelet coefficient, *w* true coefficient, and *n* noise, which is independent Gaussian. This is a classical problem in estimation theory. The MAP estimator for (2) is

$$\hat{w}(y) = \arg\max_{w} p_{w|y}(y \mid x) \tag{3}$$

Using Bayes rule, one gets

$$\begin{aligned}\hat{w}(y) &= \arg\max_{w}[p_{y|w}(y \mid w)\cdot p_w(w)] = \arg\max_{w}[p_n(y-w)\cdot p_w(w)] \\ &= \arg\max_{w}[\log(p_n(y-w)) + \log(p_w(w))]\end{aligned} \tag{4}$$

From the assumption on the noise and the wavelet coefficients, p_n is zero mean Gaussian with variance σ_n, and p_w can be described by the Laplace model which is a special case of GGD to reduce the computation complexity

$$p_w(w) = \frac{1}{\sqrt{2}\sigma}\exp\left(-\frac{\sqrt{2}\left|w\right|}{\sigma}\right) \tag{5}$$

where σ is the standard variance of the wavelet coefficients, the estimator will be

$$\hat{w}(y) = sign(y)\left(|y| - \frac{\sqrt{2}\sigma_n^2}{\sigma}\right)_+ \tag{6}$$

Since $E(n) = 0$, and w and n are statistically independent, σ in (6) is computed by $sqrt(\sigma_y^2 - \sigma_n^2)$.

4 Experimental Results

In this section, the simulation results, obtained by processing a real SAR image using our proposed method are presented and the results of the proposed approach are compared with Lee filter, Gamma MAP filter, and SWT method [9]. Fig. 1(a) depicts a 'Horse track' image (near Albuquerque, New Mexico) affected by speckle phenomena which Standard deviation to mean ratio (S/M) is equal to 0.5609. The denoised image by the proposed method is depicted in Fig. 1(d) which S/M value is reduced to 0.4255.

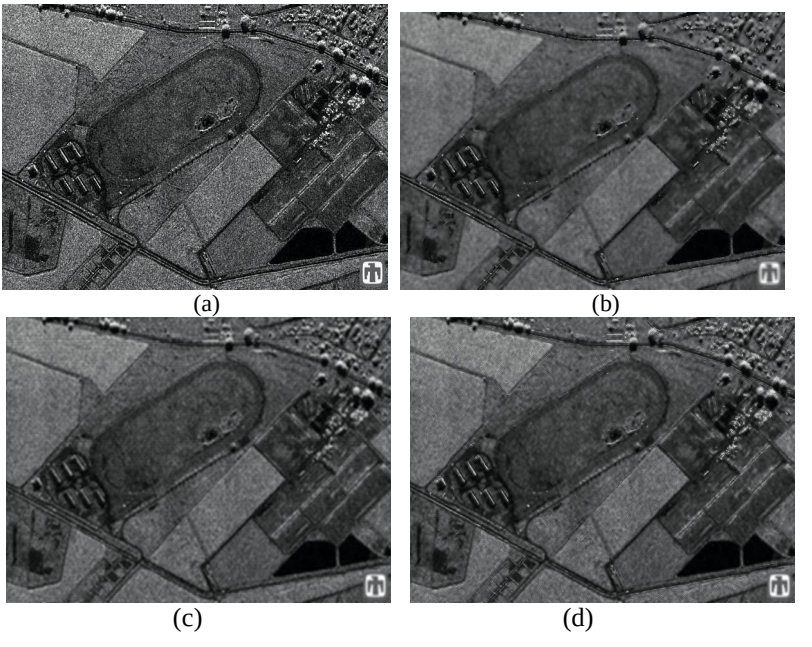


Fig. 1. Result of different despeckling methods (a) original SAR image; (b) result of Lee filter; (c) result of SWT method; (d) result of the proposed method

5 Conclusions

In this paper, we presented a wavelet-based despeckling method for SAR images. The proposed method exploits the translation invariance of an undecimated tree-structure wavelet decomposition to achieve a rejection of multiplicative noise that is practically free of the typical artifacts introduced by denoising schemes relying on critically subsampled wavelet transforms. Experimental results demonstrate that our method has a good despeckling performance in standard test and SAR images and typically preserves even the minor edges. In future work, we plan to further improve the performance of this despeckling method by including other techniques, such as the iteration process or simulated annealing.

Acknowledgment

This works was supported by Aeronautical Science Foundation of China (No. 05I53076; No. 04I53070 and) ,the National Natural Science Foundation of China (No. 60472072), and the Specialized Research Found for the Doctoral Program of Higher Education(No. 20040699034).

References

1. L. Gagnon and A. Jouan. Speckle filtering of SAR images—A comparative study between complex-wavelet-based and standard filters. Proc. SPIE, 1997.
2. F. Argenti and L. Alparone. Speckle removal from SAR images in the undecimated wavelet domain. IEEE Trans. Geosci. Remote Sensing, vol. 40, pp. 2363–2374, Nov. 2002.
3. S. Foucher, G. B. Bénié, and J.-M. Boucher. Multiscale MAP filtering of SAR images. IEEE Trans. Image Processing, vol. 10, pp. 49–60, Jan. 2001.
4. H. Guo, J. E. Odegard, M. Lang, R. A. Gopinath, I. W. Selesnick, and C. S. Burrus. Wavelet based speckle reduction with application to SAR based ATD/R. In Proc. ICIP, 1994.
5. D. L. Donoho. Denoising by soft-thresholding. IEEE Trans. Inform. Theory, vol. 41, pp. 613–627, May 1995.
6. H. Xie, L. E. Pierce, and F. T. Ulaby. Statistical properties of logarithmically transformed speckle. IEEE Trans. Geosci. Remote Sensing, vol. 40, pp. 721–727, Mar. 2002.
7. M. Mastriani, A. E. Giraldez. Smoothing of coefficients in wavelet domain for speckle reduction in Synthetic Aperture Radar images. ICGST, Vol. 6, 2005.
8. 8. L. Sendur and I. W. Selesnick. Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. IEEE Trans. Signal Processing, vol. 50, pp. 2744-2756, Nov. 2002.
9. D. Gnanadurai, V. Sadaivam. Undecimated wavelet based speckle reduction for SAR images. Pattern Recognition Letters, 26: 793-800, 2005.

An Efficient Method of Road Extraction in SAR Image

Min Wang, Yanning Zhang, and Lili Zhang

Computer school of Northwestern Polytechnical University
P.O. Box 756, Xi'an,Shaanxi, P.R. China
gaoxg@nwpu.edu.cn

Abstract. A new method of the Road extraction in synthetic aperture radar (SAR) image is proposed in this paper. Roads in a high resolution SAR image can be modeled as a homogeneous dark area bounded by two parallel lines. The fundamental is based on linear extraction and improved Hough transform. Combining Hough transform and the partition of the line, the problem of location of line segments was solved. The improved Hough transform is tested on synthetic images and noisy images. The experimental results show the method's validity.

Keywords: road extraction, Hough Transform, SAR image.

1 Introduction

Because of the speckled noise in SAR image, the structures of the Road are hidden partly. A noteworthy characteristic of the Hough Transform [1] is that it does not require the connectedness of co-linear points. Thus Hough Transform is very suitable for Road extraction from high resolution SAR images [2].

The Hough transform (HT) integrates binarization image along each possible line in the image and maps this information into a feature space. The mainly drawback of HT is that it can't get the length and the endpoints position of the line segments. To avoid the this, Zhang and Webber[3] proposed WCHT which can extract short straight lines by using small overlapping windows. However, the WCHT may extract many false short lines if the window is set to a small size in the extraction of the short lines. Y. Furukawa and Y. Shinagawa [4] determined the endpoints of the line segments by analyzing the distribution of the butterfly.

A new fast line extraction method is proposed in this paper. In this method a line segment is divided into many segments with different length which can exact location of the line. In SAR images road extraction is based on linear detection. While traditional methods need filter firstly, this decreased the singularity of the road. The method based on the improved Hough transform can solve the problem. It needn't the filter of SAR image and can get over the affection of speckle noise.

2 The Improved Hough Transform

To a pair of parameter (ρ_i, θ_i), the number of the pixels of the line segment and the position of each pixel can be gotten by Bresenham line-drawing algorithm [5]. In this

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 710 – 713, 2006.

paper, the number of pixels in a line segment is noted as $M_i(i=1,2,\cdots,n;n\geq 1)$. Form the definition of HT, we know the value of the peak in parameter space represents the number of the pixels of the white line in a black background image. So that is the same of the process of using Bresenham line-drawing algorithm, the value that is corresponded to the parameter (ρ_i,θ_i) in HT is noted as $N_i(i=1,2,\cdots,n;n\geq 1)$.

Suppose an image I with white lines on a black background and the size of the image is, $N\times N, N=2^k, k$ is integer, the improved Hough transform procedure as follow:

i. Set the original value of the layer=0, then the presently calculative image C is original image I. And suppose there is a new $N\times N$ image R in which the value of every points are zero, that is, the whole image is black.

ii. In image C, according to a group of parameters (ρ_i,θ_i) and the peak value $N_i(i=1,2,\cdots,n;n\geq 1)$, lines can be found by Hough transform.

iii. The number $M_i(i=1,2,\cdots,n;n\geq 1)$ of pixels in the line (ρ_i,θ_i) is calculated by Bresenham line-drawing algorithm.

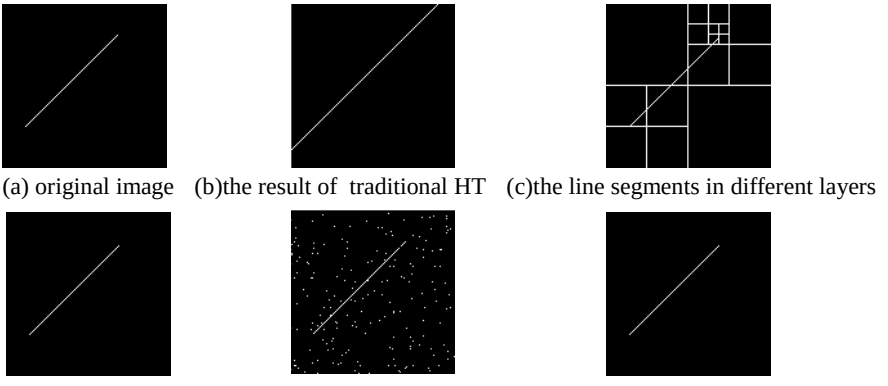
iv. Calculating the values of $\frac{N_i}{M_i}(i=1,2,\cdots,n;n\geq 1)$, if $\frac{N_k}{M_k}\geq thresh(0<thresh\leq 1)$ then the following steps will be carried out:

   a) In image R, plotting the line with the parameter (ρ_k,θ_k) run through the whole calculated range to the image C;
   b) Finding the location of the points of the line in the image C and Setting the values of these points with 0.

If $\frac{N_k}{M_k}<thresh$, then there is a line segment too which only lies in segmental range and the current layer isn't suitable for detecting the line segment. So it is necessary to keep on dividing the image by increasing the value of the layer by 1. The image C is divided into four smaller image $C_m, m=1,2,3,4$. For each C_m, supposing image C_m is image C and repeating the procedure from step ii to iv till there don't exist parameter (ρ_k,θ_k) which can satisfy the condition of $\frac{N_k}{M_k}<thresh$.

v. So the image R is the result of the extraction of the lines.

A set of 128×128 synthetic images with noise was used to test the improved Hough transform as is shown in Fig.1. There is a line segments in Synthetic image (a). Fig.1(c) shows the layers of the improved Hough transform. Comparing Fig.1(b) and Fig.1(d), the improved Hough transform can exactly locate the line while the traditional Hough transform only can get a line running though the whole image. The salt noise is added to



(a) original image (b)the result of traditional HT (c)the line segments in different layers

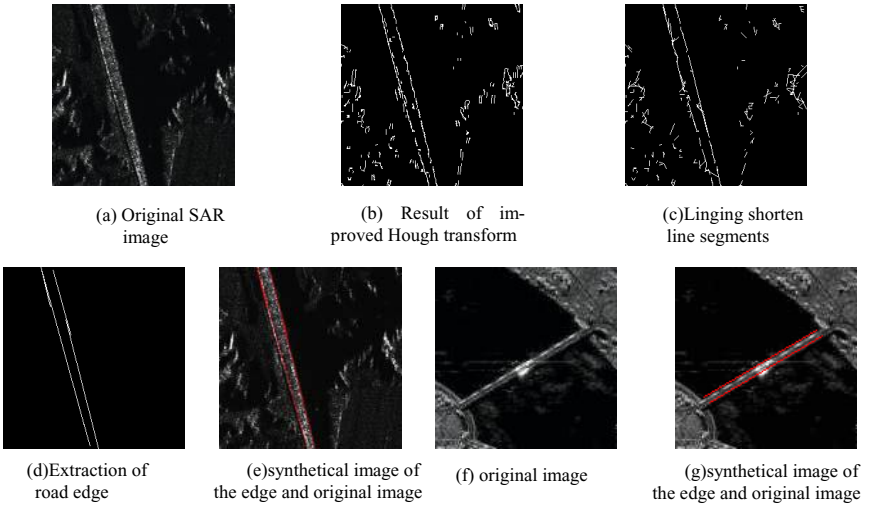
(d) the result of improved HT (e) noisy image (f)the result of improved HT to noisy image

Fig. 1. The result of the improved HT and the Comparison with traditional HT

the Fig.2(a). In Fig.2(d), the line segment still is found (see Fig.2(f)). So the method of improved Hough transform has the capability of anti-noise.

3 Experimental Results and Analysis

The first step of traditional methods of road extraction is filter while there isn't a very perfect filter method for SAR images. The processing of filter always brings future trouble to edge detection. In our method, the process of filter is thrown off because of



(a) Original SAR image (b) Result of improved Hough transform (c)Linging shorten line segments

(d)Extraction of road edge (e)synthetical image of the edge and original image (f) original image (g)synthetical image of the edge and original image

Fig. 2. Road extraction in SAR image

the ability of anti-noise of the improved Hough transform. The proposed algorithm of road extraction was tested on real SAR Images. The mainly steps are as follows.

i. Edge detection by Canny operator;
ii. Finding the short line segments by the improved Hough transform
iii. Linking the short line segments[6,7];
iv. The two edges of the road are parallel, and the distance of the two lines is more short that their lengths. So the other interferential lines in the edge image can be thrown off.

Fig.2 proved that the improved Hough transform has better effect on the roads extraction in SAR images. Based on this improved Hough transform, it also can detect bridges as is shown in Fig.2.(f),(g)

4 Conclusions

Improved Hough transform combing Hough transform and the partition of the line can exactly locate and extract lines segments from noisy images which also can be used in locating different polygons. The value of the *thresh* has influence on the accuracy of detection. We would go on the research of improving the accuracy of detection and the application in the processing of SAR images.

Acknowledgments

This works was supported by the National Natural Science Foundation of China (No. 60472072), and the Specialized Research Found for the Doctoral Program of Higher Education (No. 20040699034), and Aeronautical Science Foundation of China (No. 04153070; No.05153076), and the Natural Science Foundation of Shaan'xi Province, and Doctorate Foundation of Northwestern Polytechnical University.

References

1. J.Illingworth and A.K.Jain.: A survey of the Hough transform. Computer Vision Graphics Image Processing, vol.44.(1988)87-116.
2. JIA Cheng-li, JI Ke-feng, JIANG Yong-mei, KUANG Gang-yao.: Road Extraction from High-Resolution SAR Imagery Using Hough Transform. Geosciences and Remote Sensing Symposium, 2005. 1(2005)336 – 339.
3. Zhang Y, Webber R.: A windowing approach to detecting line segments using Hough transform, Pattern Recognition, 29:(1996)255-265.
4. Y. Furukawa and Y. Shinagawa.: Accurate and robust line segment extraction by analyzing distribution around peaks in Hough space. Computer Vision and Image Understanding, October, 92(1): (2003)1–25.
5. Jack E. Bresenham.: Algorithm for Computer Control of a Digital Plotter, IBM Systems Journal, 4(1): (1965)25-30.
6. Wessel, B.and Wiedemann, C.: The role of context for road extraction from SAR imagery. Geoscience and Remote Sensing Symposium, 6 :(2003).4025-4027.
7. **T**upin, F. and Houshmand, B.: Road detection in dense urban areas using SAR imagery and the usefulness of multiple views. IEEE Transactions on Geosciences and Remote Sensing, 40(2002):2405 – 2414.

A Novel Method for Solving the Shape from Shading (SFS) Problem

Yi Liao and Rong-chun Zhao

Northwestern Polytechnical University, Xi'an, Shaanxi, 710072, China
nwpu_liaoyi@yahoo.com, rczhao@nwpu.edu.cn

Abstract. We consider the generalized regularization problem of Shape-from-Shading. The traditional algorithms are to find the minimum point of the optimization problem where the regularization term is considered as the part of the objective function. However, the result of regularization may deviate from the true surface, due to the ill-posedness of the SFS problem. In this paper, we propose a novel method to solve this problem. The algorithm consists of two steps. In the first step, we recover the components of the surface in the range space of the transpose of the system matrix, from the observed image by using the Landweber iteration method, where the Pentland's linear SFS model is adopted without any regularization. In the second step, we represent the regularization condition as an energy spline in the Fourier domain, and find the minimum value of the energy function with respect to the components of the surface in the null space of the system matrix. Quantitative and visual comparisons, using simulated data of a fractal and smooth surface, show that the proposed method significantly outperforms the Horn, Zheng-Chellappa, Tsai-Shah and Pentland linear methods for surface reconstruction.

1 Introduction

Shape-from-Shading (SFS) has been researched for many years and applied in different areas. Although recent work on SFS claims that it can be modeled as a well-posed problem by assuming a pinhole camera and a point light source located at the optical center [1], [2], [3], in other situations SFS problem is ill-posed [8], [10] and the variety of regularization methods [4], [5], [6], [7], [23], [24] have been applied to obtain a stable result. Generally the regularization term is considered as the part of the objective function to optimize but the result may deviate from the true surface [4], [5], [6], [7]. However, Worthington and Hancock provided a new framework where the image irradiance equation is a hard constraint and the needle-map consistency constraint is used for updating the needle-map [23]. This improves the data closeness of the recovered needle-map and facilitates control of the new framework to investigate various types of needle-map consistency constraint. In this paper we propose a new method to keep the data closeness and a one-step update by using a generalized regularizer. The proposed method recovers the surface height directly rather than from a needle-map of the surface.

The discrete surface and image are considered where the surface size is larger than the image size. Using the Pentland's linear SFS model [9], we first reconstruct the

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 714 – 723, 2006.

surface by using the Landweber iteration method. And then we provide a null space compensation method to recover the components of the surface. This is in the null space of the system matrix and cannot be obtained from the Landweber iteration. The regularization term is minimized with respect to the null space components of the surface to be reconstructed, where the regularization term is formed as a spline represented as an energy function in the Fourier domain [16].

In section 2, we introduce and apply the Landweber iteration to the Pentland's linear SFS model to recover the range space components of the surface that is to be reconstructed. In section 3, we propose our null space compensation method for LSFS problem, with the regularization term represented as a variational spline in Fourier domain. Simulation results are shown in section 4, followed by conclusions in section 5.

2 Linear SFS Based on Landweber Iteration

The fundamental model of the SFS problem [4] is

$$f(x,y)=\frac{ap(x,y)+bq(x,y)+c}{\sqrt{p^2(x,y)+q^2(x,y)+1}} \tag{1}$$

where a,b,c are the parameters determined by the illumination direction; $f(x,y)$ is the image intensity; $p=\partial z/\partial x$, $q=\partial z/\partial y$ with $z(x,y)$ being the surface height corresponding to image $f(x,y)$. When the surface slope is low, i.e. $\sqrt{1+p^2+q^2}\approx 1$, Pentland's linear model [9] (LSFS) may be used:

$$f(x,y)\cong ap(x,y)+bq(x,y)+c\,. \tag{2}$$

We denote the LSFS problem as a general matrix problem

$$\boldsymbol{Az}=\boldsymbol{f} \tag{3}$$

where $\boldsymbol{A}$ is a coefficient matrix determined by equation (2), $\boldsymbol{z}$ a vector representing the surface height to be recovered, $\boldsymbol{f}$ a vector which satisfies $\boldsymbol{f}=\hat{\boldsymbol{f}}+\boldsymbol{c}$. $\hat{\boldsymbol{f}}$ is the vector form of the observed image and $\boldsymbol{c}$ is a constant vector caused by linearization.

The basic Landweber iteration to solve problem (3) is as follows [12]:

$$\boldsymbol{z}^{(n+1)}=\boldsymbol{z}^{(n)}-\tau\boldsymbol{A}^T\left(\boldsymbol{Az}^{(n)}-\boldsymbol{f}\right) \tag{4}$$

where τ is a constant that has to meet $0<\tau<1/\|\boldsymbol{A}\|^2$. Many modified Landweber iteration methods can be found in different applications, such as recent developments on image reconstruction in Electrical Impedance Tomography [12], [13], [14], [15]. However, these modifications do not perform much better than the basic Landweber iteration method according to Pan and Yagle [11]. They derived that the k-th iteration can be represented as

$$\mathbf{z}_{k+1} = \mathbf{z}_0^{\perp} + \sum_{i=1}^{r(\mathbf{A})}\left(1-\left(1-\sigma_i^2\right)^k\right)\frac{1}{\sigma_i}\bar{\mathbf{f}}^T\mathbf{u}_i\mathbf{v}_i + \sum_{i=1}^{r(\mathbf{A})}\left(1-\left(1-\sigma_i^2\right)^k\right)\frac{1}{\sigma_i}\boldsymbol{\varepsilon}^T\mathbf{u}_i\mathbf{v}_i + \sum_{i=1}^{r(\mathbf{A})}\left(1-\sigma_i^2\right)^k\tilde{\mathbf{z}}_0^T\mathbf{u}_i\mathbf{v}_i \tag{5}$$

where the orthogonal matrices $\boldsymbol{U}=[\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m], \boldsymbol{V}=[\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n]$ and the diagonal matrix $\boldsymbol{S}$ form the singular value decomposition (SVD) $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$. σ_i is the singular values of matrix $\boldsymbol{A}$. $r(\boldsymbol{A})$ is the rank of $\boldsymbol{A}$ and

$$\tilde{\boldsymbol{z}}_0 = \sum_{i=1}^{r(\boldsymbol{A})} w_i\boldsymbol{v}_i \ , \ \boldsymbol{z}_0^{\perp} = \sum_{i=1}^{n-r(\boldsymbol{A})} w_{i+r(\boldsymbol{A})}\boldsymbol{v}_{i+r(\boldsymbol{A})} \tag{6}$$

are the components of the initial value $\boldsymbol{z}_0=\tilde{\boldsymbol{z}}_0+\boldsymbol{z}_0^{\perp}$ in the range space $R(\boldsymbol{A}^T)$ and the null space $N(\boldsymbol{A})$, respectively. $w_i, i=1,\cdots,n$ are the coefficients. The first, second and third sums in the above equation represent the reconstruction from the data, the error in the observed data, and the initial value, respectively. As $k\to\infty$, the first and second sum become constants, and the third sum approaches zero. It means that the effect of the components of initial value in $R(\boldsymbol{A}^T)$ die out, the data will be reconstructed, but at the same time the error will be amplified. This explains the deterioration problem in [13].

3 Null Space Compensation with an Energy Spline

From section 2 we can see that the components of the surface $\boldsymbol{z}$ in the null space of the matrix $\boldsymbol{A}$ cannot be reconstructed by the Landweber Iteration method (or any generalized Landweber iteration method [11]). However, the null space components may be very important to the reconstructed surface if the rank $r(\boldsymbol{A})$ is relatively low. This means that lots of information of $\boldsymbol{z}$ is lost during the transformation and cannot be recovered without any other information on $\boldsymbol{z}$. In this section, we propose a novel method to compensate the null space components of terrain surface $\boldsymbol{z}$ with a spline represented as an energy function in the Fourier domain [16], based on the reconstructed result from the Landweber iteration method.

Suppose the surface $\boldsymbol{z}$ can be represented by the orthogonal matrix $\boldsymbol{V}$ as

$$\boldsymbol{z}=\boldsymbol{V}\boldsymbol{w} \tag{7}$$

where $\boldsymbol{w}$ is the coefficient vector that denotes the projection of $\boldsymbol{z}$ onto $\boldsymbol{V}$:

$$\boldsymbol{w}=\boldsymbol{V}^T\boldsymbol{z}\,. \tag{8}$$

Without loss of generality, we assume that $\boldsymbol{w}$ is such that

$$\boldsymbol{w}=\begin{pmatrix}\boldsymbol{w}^{\perp}\\ \tilde{\boldsymbol{w}}\end{pmatrix}=\begin{pmatrix}\boldsymbol{V}_{w^{\perp}}^T\\ \boldsymbol{V}_{\tilde{w}}^T\end{pmatrix}\boldsymbol{z} \tag{9}$$

where $\boldsymbol{V}_{\boldsymbol{w}^{\perp}}, \boldsymbol{V}_{\tilde{\boldsymbol{w}}}$ denote the sub-matrices consisting of the basis corresponding to the null space components $\boldsymbol{w}^{\perp}$ and the range space components $\tilde{\boldsymbol{w}}$ respectively.

Given the expectation $E[\boldsymbol{z}]=\boldsymbol{\theta}$ and the covariance matrix $\boldsymbol{C}$ of $\boldsymbol{z}$, the expectation and covariance matrix of $\boldsymbol{w}$ are $\boldsymbol{\mu}=\boldsymbol{V}^T\boldsymbol{\theta}$ and $\boldsymbol{V}^T\boldsymbol{C}\boldsymbol{V}$ respectively. We infer the unknown components of the surface according to the given statistics, i.e. sample the unknown components with highest probability:

$$\max_{\boldsymbol{w}^{\perp}} \pi\left(\boldsymbol{w}^{\perp}\middle|\tilde{\boldsymbol{w}}\right)=\max_{\boldsymbol{w}^{\perp}} \exp\left(-\frac{1}{2}(\boldsymbol{w}-\boldsymbol{\mu})^T\left(\boldsymbol{V}^T\boldsymbol{C}\boldsymbol{V}\right)^{-1}(\boldsymbol{w}-\boldsymbol{\mu})\right),$$

where π denotes the probability function. It is equivalent to solving the following minimization problem:

$$\min_{\boldsymbol{w}^{\perp}}\left\{\frac{1}{2}\begin{pmatrix}\boldsymbol{w}^{\perp}-\boldsymbol{\mu}^{\perp}\\ \tilde{\boldsymbol{w}}-\tilde{\boldsymbol{\mu}}\end{pmatrix}^T\begin{pmatrix}\boldsymbol{V}_{\boldsymbol{w}^{\perp}}^T\boldsymbol{C}^{-1}\boldsymbol{V}_{\boldsymbol{w}^{\perp}} & \boldsymbol{V}_{\boldsymbol{w}^{\perp}}^T\boldsymbol{C}^{-1}\boldsymbol{V}_{\tilde{\boldsymbol{w}}}\\ \boldsymbol{V}_{\tilde{\boldsymbol{w}}}^T\boldsymbol{C}^{-1}\boldsymbol{V}_{\boldsymbol{w}^{\perp}} & \boldsymbol{V}_{\tilde{\boldsymbol{w}}}^T\boldsymbol{C}^{-1}\boldsymbol{V}_{\tilde{\boldsymbol{w}}}\end{pmatrix}\begin{pmatrix}\boldsymbol{w}^{\perp}-\boldsymbol{\mu}^{\perp}\\ \tilde{\boldsymbol{w}}-\tilde{\boldsymbol{\mu}}\end{pmatrix}\right\} \tag{10}$$

where $\boldsymbol{\mu}^{\perp}, \tilde{\boldsymbol{\mu}}$ are the components of $\boldsymbol{\mu}$ in the null space $N(\boldsymbol{A})$ and the range space $R(\boldsymbol{A}^T)$ respectively. The minimum value of equation (10) has to satisfy

$$\boldsymbol{C}_{\boldsymbol{w}^{\perp}}^{inv}\left(\boldsymbol{w}^{\perp}-\boldsymbol{\mu}^{\perp}\right)+\boldsymbol{C}_{\boldsymbol{w}^{\perp},\tilde{\boldsymbol{w}}}^{inv}\left(\tilde{\boldsymbol{w}}-\tilde{\boldsymbol{\mu}}\right)=\boldsymbol{0} \tag{11}$$

where

$$\boldsymbol{C}_{\boldsymbol{w}^{\perp}}^{inv}=\boldsymbol{V}_{\boldsymbol{w}^{\perp}}^T\boldsymbol{C}^{-1}\boldsymbol{V}_{\boldsymbol{w}^{\perp}},\ \boldsymbol{C}_{\boldsymbol{w}^{\perp},\tilde{\boldsymbol{w}}}^{inv}=\boldsymbol{V}_{\boldsymbol{w}^{\perp}}^T\boldsymbol{C}^{-1}\boldsymbol{V}_{\tilde{\boldsymbol{w}}}.$$

More specifically, we provide the following proposition for the preferred structure of the prior information of the surface to be reconstructed.

Proposition 1 (necessary condition): Let $\boldsymbol{C}^{-1}=\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T$ and $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$ be the singular value decomposition of the symmetric matrix $\boldsymbol{C}^{-1}$ and the system matrix $\boldsymbol{A}$ respectively. Suppose $\boldsymbol{V}$ satisfies the equation (9). If $\boldsymbol{C}_{\boldsymbol{w}^{\perp},\tilde{\boldsymbol{w}}}^{inv}=\boldsymbol{V}_{\boldsymbol{w}^{\perp}}^T\boldsymbol{C}^{-1}\boldsymbol{V}_{\tilde{\boldsymbol{w}}}$ is not a zero matrix, then the inverse $\boldsymbol{C}^{-1}$ satisfies:

(1) The diagonal matrix $\boldsymbol{\Lambda}\neq\alpha\boldsymbol{I}$, $\boldsymbol{\Lambda}\neq\boldsymbol{O}$; and

(2) The orthogonal matrix $\boldsymbol{P}$ is not parallel to the orthogonal matrix $\boldsymbol{V}$. Here parallel means that $\boldsymbol{P}^T\boldsymbol{V}=diag(\{1,-1\})$, where $diag(\{1,-1\})$ is a diagonal matrix with each diagonal element being either 1 or -1.

Proof: We first show that the first condition has to hold. Obviously, if $\boldsymbol{\Lambda}=\boldsymbol{O}$, then $\boldsymbol{C}_{\boldsymbol{w}^{\perp},\tilde{\boldsymbol{w}}}^{inv}=\boldsymbol{V}_{\boldsymbol{w}^{\perp}}^T\boldsymbol{C}^{-1}\boldsymbol{V}_{\tilde{\boldsymbol{w}}}=\boldsymbol{V}_{\boldsymbol{w}^{\perp}}^T\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T\boldsymbol{V}_{\tilde{\boldsymbol{w}}}=\boldsymbol{O}$ is a zero matrix. If $\boldsymbol{\Lambda}=\alpha\boldsymbol{I}$, then

$$\boldsymbol{C}_{\boldsymbol{w}^{\perp},\tilde{\boldsymbol{w}}}^{inv}=\boldsymbol{V}_{\boldsymbol{w}^{\perp}}^T\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T\boldsymbol{V}_{\tilde{\boldsymbol{w}}}=\alpha\boldsymbol{V}_{\boldsymbol{w}^{\perp}}^T\boldsymbol{P}\boldsymbol{P}^T\boldsymbol{V}_{\tilde{\boldsymbol{w}}}=\boldsymbol{O}.$$

Now we consider the second condition. $\forall\boldsymbol{v}_j\in\boldsymbol{V}_{\tilde{\boldsymbol{w}}}=(\boldsymbol{v}_{n-r(\boldsymbol{A})+1},\cdots,\boldsymbol{v}_n)$, its projection on $\boldsymbol{P}$ is

$$\boldsymbol{P}^T\boldsymbol{v}_j = \left(\underbrace{0,\cdots,0,\pm 1}_{j},\cdots,0\right)^T.$$

Then

$$\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T\boldsymbol{v}_j = \boldsymbol{P}\left(\underbrace{0,\cdots,0,\pm\gamma_j}_{j},\cdots,0\right)^T = \pm\gamma_j\boldsymbol{p}_j,$$

where $\boldsymbol{p}_j \in \boldsymbol{P}_{\tilde{\boldsymbol{w}}} = \left(\boldsymbol{p}_{n-r(\boldsymbol{A})+1},\cdots,\boldsymbol{p}_n\right)$. If $\boldsymbol{P}^T\boldsymbol{V} = diag(\{1,-1\})$, then we have

$$\boldsymbol{v}_i^T\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T\boldsymbol{v}_j = \pm\gamma_j\boldsymbol{v}_i^T\boldsymbol{p}_j = 0,\ \forall\boldsymbol{v}_i \in \boldsymbol{V}_{\boldsymbol{w}^\perp} = \left(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{n-r(\boldsymbol{A})}\right).$$

Thus

$$\boldsymbol{C}^{inv}_{\boldsymbol{w}^\perp,\tilde{\boldsymbol{w}}} = \boldsymbol{V}^T_{\boldsymbol{w}^\perp}\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T\boldsymbol{V}_{\tilde{\boldsymbol{w}}} = \boldsymbol{O}.$$

□

We seek the orthogonal matrix $\boldsymbol{P}$ such that the projection of each component of $\boldsymbol{V}$ onto any axis of $\boldsymbol{P}$ is nonzero. Szeliski and Terzopoulos [16] provided a regularization term for surface reconstruction problem, where the regularization term is a spline represented as an energy function:

$$\frac{1}{2}\sum_{m=0}^{n}\iint\mu_m(x,y)\sum_{k+l=m}\frac{m!}{k!l!}\left|\frac{\partial^m z(x,y)}{\partial x^k\partial y^l}\right|^2 dxdy$$

where $\mu_m(x,y)$ is the weight. The Fourier transformation of the above functions when μ_m are constant is $\frac{1}{2}\int|G(\boldsymbol{\omega})|^2|\hat{z}(\boldsymbol{\omega})|^2 d\boldsymbol{\omega}$, where $\hat{z}(\boldsymbol{\omega})$ is the Fourier transform of the surface $z(x,y)$ and $|G(\boldsymbol{\omega})|^2 = \sum_{m=0}^{n}\mu_m|\boldsymbol{\omega}|^{2m}$. If $z(x,y)$ is a random field, then $\hat{z}(\boldsymbol{\omega})$ is a Gaussian variable with variance $|G(\boldsymbol{\omega})|^{-2}$. Szeliski and Terzopoulos extend the variance as $|\boldsymbol{\omega}|^{-\beta}$ by comparing it with the power spectrum of a fractal Brownian surface. Thus the prior distribution of $\boldsymbol{z}$ can be represented as:

$$\pi(\boldsymbol{z}) \propto \exp\left(-\frac{1}{2}\left(\boldsymbol{P}^T\boldsymbol{z}\right)^T\boldsymbol{\Lambda}\left(\boldsymbol{P}^T\boldsymbol{z}\right)\right) \tag{12}$$

where $\boldsymbol{P}$ denotes a Fourier matrix. Covariance matrix $\boldsymbol{\Lambda}^{-1}$ is diagonal with each diagonal element being $|\boldsymbol{\omega}|^{-\beta}$ that corresponds to the same indices $\boldsymbol{\omega}$ in $\boldsymbol{P}$. The exponent $2 \le \beta \le 4$. When $\beta = 4$, it refers to a relatively smooth surface. When $2 < \beta < 4$, it refers to a fractal surface.

The equation (11) can be solved by any matrix problem solver [20]. We describe our null space compensation algorithm as follows.

Algorithm (Null Space Compensation)**:** Given problem (3) and the prior information $\boldsymbol{C}^{-1} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T$ and supposing $\boldsymbol{\mu} = \boldsymbol{0}$ (this can be down by extracting the mean value):

(a) Do the normal Landweber iteration using equation (4) until convergent with given initial value $\boldsymbol{z}_0$, obtain the reconstructed data $\boldsymbol{z}$;

(b) Project $\boldsymbol{z}$ and $\boldsymbol{z}_0$ onto the orthogonal matrix $\boldsymbol{V}$ using equation (8), obtain the coefficient vector $\boldsymbol{w}$ and $\boldsymbol{w}_0$ respectively;

(c) Find the invariant components in $\boldsymbol{w}$ and $\boldsymbol{w}_0$, obtain the components of $\boldsymbol{z}$ and $\boldsymbol{z}_0$ in the null space $N(\boldsymbol{A})$, denote it as $\boldsymbol{w}_0^{\perp}$;

(d) Subtract the invariant components $\boldsymbol{w}_0^{\perp}$ from $\boldsymbol{w}$ and obtain the components of $\boldsymbol{z}$ in the range space $R(\boldsymbol{A}^T)$, denote it as $\tilde{\boldsymbol{w}}$;

(e) Compute the new components $\boldsymbol{w}^{\perp}$ of the reconstructed data in null space $N(\boldsymbol{A})$, by minimizing $\left(\boldsymbol{P}^T\boldsymbol{V}\boldsymbol{w}^*\right)^T \boldsymbol{\Lambda}\left(\boldsymbol{P}^T\boldsymbol{V}\boldsymbol{w}^*\right)$ with respect to $\boldsymbol{w}^{\perp}$, where $\boldsymbol{w}^* = \left(\left(\boldsymbol{w}^{\perp}\right)^T, \tilde{\boldsymbol{w}}^T\right)^T$.

(f) Compute the new reconstructed data $\boldsymbol{z}^* = \boldsymbol{V}\boldsymbol{w}^*$.

The complexity of the algorithm mainly depends on the SVD operator which is $O(N^2)$. Thus the computation time increases dramatically when the image size increases. Its applicability depends on the performance of the hardware. However, in practice we can always apply multi-analysis [21] to a large-scale problem or recover the surface patch by patch, which results small-scale problems.

4 Simulations

In this section, we describe the simulation of a small fractal surface of size 17×17 with fractal dimension 2.2 (see Fig. 1 (a)) by the Power Spectral Synthesis method [17], and a surface of a semi sphere of size 17×17(see Fig. 1 (b)). The corresponding simulated images, of size 16×16 with illumination (45^o, 45^o), are presented in figures 1(1)

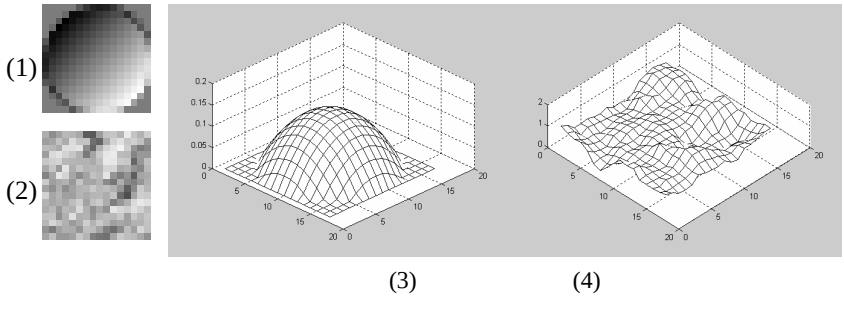


Fig. 1. Simulated Surfaces and Images. (3) is a surface of a semi sphere and (4) is a fractal surface. (1) and (2) are images generated with illumination direction as (45^o, 45^o) from the ground truth (3) and (4) respectively.

and 1(2) based on equation (1) respectively. Note that the box dimension [18] of the simulated fractal surface is not exactly 2.2, but 2.2541. Simulated data provides a valid and easily reproducible basis for *comparative* analysis of the proposed method with other known methods.

In the implementation of our method, we set $\beta = 2\times 1.0 + 2 = 4$ for the image (1) and $\beta = 2\times 0.8 + 2 = 3.6$ for the image (2) in Fig. 1, since the former one is smooth and the latter one is fractal. Alternatively, we can use the method provided by Pentland to estimate the dimension of a fractal surface from its image intensity [22]. Set the parameter $\tau = 0.5$ in the Landweber iteration. Before computing the squared error between the reconstructed

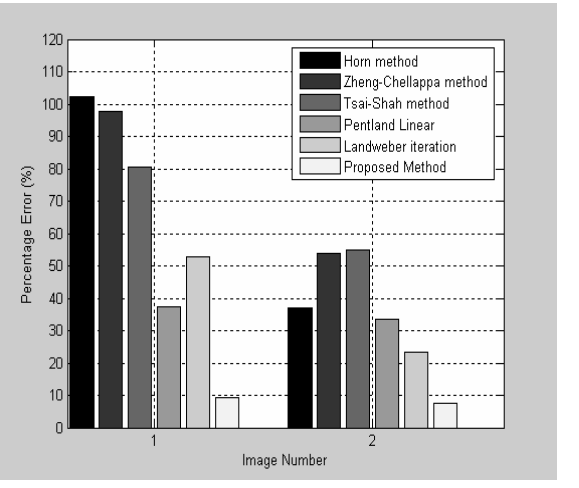
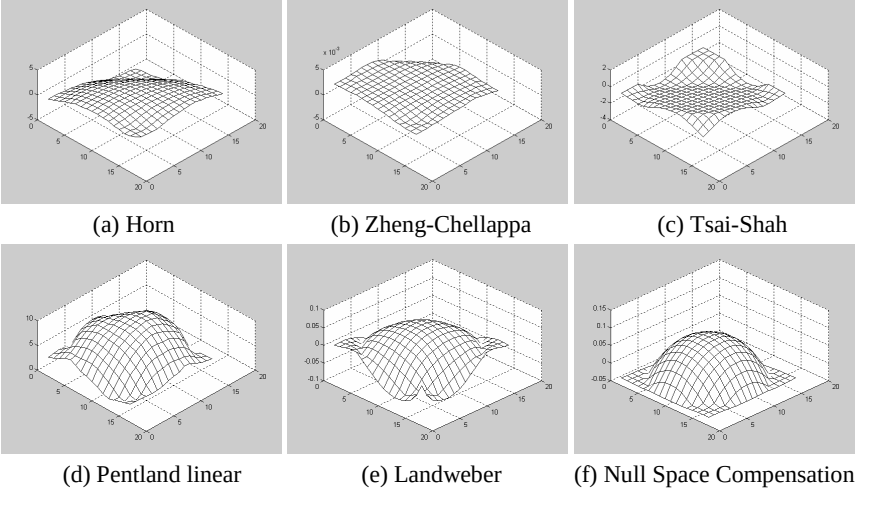


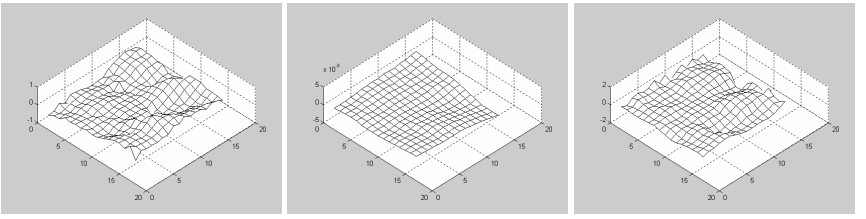
Fig. 2. Comparison of the Percentage Errors. In this figure we compare the percentage error of our proposed method with the Landweber iteration and four typical Shape-From-Shading methods for the images (1) and (2) in Fig. 1. The proposed method significantly outperforms the other methods in both smooth (corresponding to image (1)) and fractal (corresponding to image (2)) surfaces.



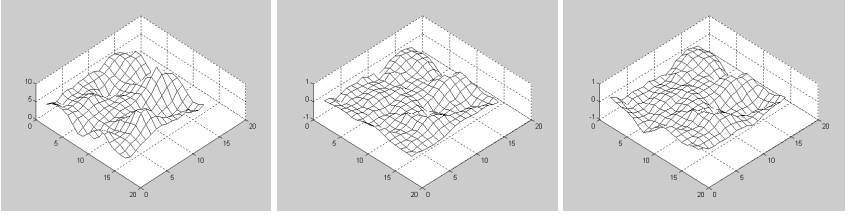
(a) Horn (b) Zheng-Chellappa (c) Tsai-Shah

(d) Pentland linear (e) Landweber (f) Null Space Compensation

Fig. 3. The Reconstructed Smooth Surfaces. In this figure the reconstructed surfaces for image (1) in Fig. 1 from typical SFS algorithms, the Landweber iteration, and our null space compensation method are shown. The proposed method is a significantly more accurate reconstruction than the other known methods.



(a) Horn (b) Zheng-Chellappa (c) Tsai-Shah



(d) Pentland linear (e) Landweber (f) Null Space Compensation

Fig. 4. The Reconstructed fractal Surfaces. In this figure the reconstructed surfaces for image (2) in Fig. 1 from typical SFS algorithms, the Landweber iteration, and our proposed method are shown. At first sight, the Horn, Landweber and proposed method are close to the model surface, however, a more detailed visual comparison shows the proposed method to be more accurate.

surface and the ground truth, we force the value of data between 0 and 1 and subtract the mean value from the normalized data as $\hat{z}_{i,j} = \tilde{z}_{i,j} - \mu$, where

$$\tilde{z}_{i,j} = \frac{z_{i,j} - s}{t - s}, \quad s = \min\{z_{i,j}\}, \quad t = \max\{z_{i,j}\}, \quad \mu = \frac{1}{total}\sum_{i,j}\tilde{z}_{i,j}.$$

We compare the results from four commonly used SFS algorithms with our method. The four SFS methods selected were Horn's method [4], the Zheng-Chellappa method [5], the Tsai-Shah method [7], and Pentland's linear method [9]. A survey [19] has been conducted on these methods (the source code is available from http: //www.cs.ucf.edu/ ~vision/ source.html). The results from these methods are compared with the Landweber iteration method and our null space compensation method in terms of percentage error between the reconstructed surface and the ground truth in Fig. 1 (3) and (4) defined as:

$$percentage\_err = \sqrt{\sum_{i,j}\left(z_{i,j} - z^*_{i,j}\right)^2 \Big/ \sum_{i,j}\left(z^*_{i,j}\right)^2} \times 100\%$$

where z^* denotes the ground truth. They are shown in Fig. 2. The reconstructed surfaces from images (1) and (2) in Fig. 1 are compared in Fig. 3 and Fig. 4 respectively. We can see that the percentage error of the reconstructed surfaces from our null space compensation method is much lower (below 10%) compared with other methods in both smooth and fractal surface cases. The Landweber iteration method can recover the surface to some extent and works even better than the Pentland's linear

method when the Pentland's linear SFS model is suitable, e.g. image (2) of a fractal surface (4) in Fig. 1. On the other hand, it suffers from the errors caused by linearization when linear SFS model is not suitable, and performs worse than the Pentland's linear method. However, when using our method to compensate for the components that cannot be recovered by the Landweber iteration method, the error in the reconstructed surface compared with the ground truth dramatically decreases. It outperforms the usual regularization methods such as the Horn's method, the Zheng-Chellappa method, and the Tsai-shah method.

5 Conclusions

In this paper, we provide a new method for the generalized regularization problem of shape from shading (SFS). The Pentland's linear SFS model is adopted. By assuming a larger surface size than the image size from a finite difference discretization and representing the linear SFS problem as a general ill-posed matrix problem, we proposed a null space compensation method based on the Landweber iteration with a Fourier-based energy function to solve it. The novel idea is that we minimize the energy function with respect to the components of the surface in the null space of the system matrix, based on the components reconstructed by the Landweber iteration method. The experimental results on the small-scale simulated data strongly support our method. The comparisons between our method and four typical SFS methods, the Horn, Zheng-Chellapa, Tsai-Shah and Pentland linear, show that the proposed method significantly outperforms other methods in both smooth and fractal surfaces' cases, although the nonlinear part in the image intensity is not removed and considered as the noise. Future work can focus on large-scale problems, which are more common in practice.

Acknowledgement

The authors would like to thank Professor Maria Petrou (Imperial College, UK) for kindly reviewing the paper. Yi Liao would like to thank Munir Zaman (University of Surrey, UK) for his useful comments on the structure of the paper.

References

1. Prados , E., Faugeras , O.: Shape From Shading: A Well-Posed Problem?. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, California, Volume II (2005) 870–877
2. Prados , E., Faugeras, O.: Unifying Approaches and Removing Unrealistic Assumptions in Shape from Shading: Mathematics can help. In Proceedings of the European Conference on Computer Vision (ECCV'04), Prague, Czech Republic (2004)
3. Prados, E., Faugeras, O., Camilli, F.: Shape from Shading: a well-posed problem?. Technical Report RR-5297, INRIA (2004)
4. Horn, B. K. P.: Height and Gradient from Shading. International Journal of Computer Vision, 5(1), (1990) 37–75

5. Zheng, Q., Chellappa, R.: Estimation of Illuminant Direction, Albedo, and Shape from Shading. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(7), (1991) 680–702
6. Lee , C.-H., Rosenfeld, A.: Improved Methods of Estimating Shape from Shading Using the Light Source Coordinate System. Artificial Intelligence, 26 (1985) 125–143
7. Tsai, P.-S., Shah, M.: A Fast Linear Shape from Shading. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (1992) 734–736
8. Atsuta, K., Sanoworasil, M., Kondo, S.: A New Method of Estimating Shape from Shading by using a Constraint at Zero-crossings of Image Brightness. In: Proceedings of 9th International Conference on Pattern Recognition, 2(1988) 666–668
9. Pentland, A.: Shape Information from Shading: A Theory about Human Perception. In Proceedings of International Conference on Computer Vision (1988) 404–413
10. Ikeuchi, K., Horn, B. K. P.: Numerical Shape from Shading and Occluding Boundaries. Artificial Intelligence, 17 (1981) 141–184
11. Pan, T. S., Yagle, A. E.: Numerical Study of Multigrid Implementations of Some Iterative Image Reconstruction Algorithms. IEEE Transactions on Medical Imaging, 10(4), (1991) 572–588
12. Wang, H., Wang, C., Yin, W.: A Pre-Iteration Method for the Inverse Problem in Electrical Impedance Tomography. IEEE Transactions on Instrumentation and Measurement, 53(4), (2004) 1093–1096
13. Liu, S., Fu, L., Yang, W. Q., Wang, H. G., Jiang, F.: Prior-Online Iteration for Image Reconstruction with Electrical Capacitance Tomography. IEE Proceedings- Measurement Science and Technology, 151(3), (2004) 195–200
14. Zhao, J., Fu, W., Li, T., Wang, S.: An Image Reconstruction Algorithm Based on A Revised Regularization Method for Electrical Capacitance Tomography. Measurement Science and Technology, 13 (2002) 638–640
15. Wang, W. Q., Spink, D. M., York, T. A., McCann, H.: An Image-Reconstruction Algorithm Based on Landweber's Iteration Method for Electrical-Capacitance Tomography. Measurement Science and Technology, 10 (1999) 1065–1069
16. Szeliski, R., Terzopoulos, D.: From Splines to Fractals. Computer Graphics, 23(3), (1989) 51–60
17. Turner, M. J., Blackledge, J. M., Andrews, P. R.: Fractal Geometry in Digital Imaging. Academic Press (1998) 525 B Street, Suite 1900, San Diego, California 92101-4495, USA
18. Russ, J. C.: Fractal Surfaces. Plenum Press, New York and London (1994)
19. Zhang, R., Tsai, P.-S., Cryer, J. E., Shah, M.: Shape-from-Shading: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(8), (1999) 690–706
20. Barrett, R., et. al.: Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods. Society for Industrial and Applied Math (1993)
21. Blu, T., Unser, M.: Wavelets, fractals and radial basis functions. IEEE Transactions on Signal Processing, 50(3), (2002) 543–553
22. Kube, P., Pentland, A.: On the Imaging of Fractal Surfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 10(5), (1988) 704–707
23. Worthington, P. L., Hancock, E. R.: New Constraints on Data-Closeness and Needle Map Consistency for Shape-from-Shading. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(12), (1999) 1250–1267
24. Ragheb, H., Hancock, E. R.: A Probabilistic Framework for Specular Shape-from-Shading. Pattern Recognition, 36 (2003) 407–427

A New Fast Algorithm for Training Large Window Stack Filters*

Guangming Shi [1], Weisheng Dong [1], Li Zhang [1], and Jin Pan [2]

[1] School of Electronic Engineering, Xidian University, Xi'an, China, 710071
wsdong@mail.xidian.edu.cn, gmshi@xidian.edn.cn,
zhang1_li2@163.com
[2] Lab of Network Security and Countermeasure, Xi'an Communications

Abstract. Stack filters are often employed for suppressing the pulse noise. In general, the larger sizes the stack filters are, the better results are. Unfortunately, available algorithms for designing stack filters can only be suit for small window sizes due to their huge computational complexities. This paper presents a new fast adaptive algorithm for designing a stack filter with large windows. The idea of the new algorithm is to divide a lager window into many sub-windows. The procedures of dividing a large window are given. An Immune Memory Clonal Selection Algorithm is employed to design the stack filters with small window sizes. Because of its highly parallel structure, it can be very fast implemented. As an experiment, the algorithm was used to restore images corrupted by uncorrelated additive noise with the level from 10% to 50 %. The results show that the algorithm is effective and feasible.

1 Introduction

Stack filters are a new class of nonlinear, sliding-window filters that defined by a weak superposition property known as the threshold decomposition and an ordering property called the stacking properties [1]-[2]. Some well-known filters, such as the standard median [3], rank order filters [4], morphological filters [5], and all compositions of these filters can be expressed as stack filters [1], [6]. Because of their ability to adapt to different types of noise and be suit for VLSI implementation [7], they are widely used in image processing applications as image enhancement, and edge detection in noisy image [8]. The difficulty is that the number of stack filters grows very quickly as the size of the filter window increases.

A class of approaches is by making use of the adaptive training approach to deal with this hard problem and attempts to obtain a good solution in a comparatively shorter time. The adaptive approach arrives at an optimal or near-optimal stack filter through an iterative approach that relies on training data. It is time consuming and the convergence speed is slow. The FASTAF algorithm developed in [9] is often fast while designing stack filters with small window sizes, such as 3×3. Unfortunately, this good feature disappears as the window size grows to 4×4 and can only be used to design filters with up to 11-point windows. The new FASTAF algorithm developed

* Nation Natural Science Fund of China Under Grant No. 60372047.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 724 – 733, 2006.

in [10] differs in both the amount of training and the method used to enforce the stacking constraint. With this new algorithm, the stack filters with window sizes as large as 22 points can be developed. However, this algorithm will still become unmanageable while designing stack filters with larger window sizes, such as 5×5 and 7×7.

The stack filters with large windows are often needed to produce the desired visual result in image processing applications. Unfortunately, available algorithms for designing stack filters can only be suit for small window sizes due to their huge computational complexities. In our opinions, there exist unreasonable searching strategies in these algorithms. In fact, it is not necessary to perform a search over the set of stack filters. In this paper, a novel stack filter training algorithm is developed. Each filter designed in this work is in a cascade of several 3×3 stack filters, and this algorithm enables us to use very large window sizes. This new algorithm is also highly parallel in nature. It allows very fast implementations to be developed when massively parallel machines are available. In this work, the 3×3 stack filters are configured by the Immune Memory Clonal Selection Algorithm (IMCSA) proposed in [11]. Using the algorithm, the 5×5, 7×7 or larger window stack filters can be designed. The given experiments show the efficiency of this proposed method.

2 Stack Filters

Stack filters are defined by two important properties – the threshold decomposition property and the stacking property [1]-[2]. Input signal $X=(x_1,x_2,\cdots,x_n)$ is assume to be an *M* value, length-n vector, and it can be represented as the sum of a series of binary-valued sequence,

$$X(i)=\sum_{l=1}^{M}x_l(i),\ \text{where}\ X(i)=x_i, i=1,2,\cdots,n \tag{1}$$

where for each l, the binary sequence $x_l(i)$ is obtained by thresholding $X(i)$, as:

$$x_l(i)=T_l(X(i))=\begin{cases}1, & X(i)\ge l\\ 0, & X(i)<l\end{cases} \tag{2}$$

where T_l is a threshold operator. For a stack filter S, the output $S(X)$ is given by:

$$S(X)=\sum_{l=1}^{M-1}S(T_l(X)) \tag{3}$$

where $S(T_l(X))$ is the output of the filter S operating on a binary signal $T_l(X)$.

Each stack filter is defined by a positive Boolean function (PBF) $f(\cdot)$ which can be uniquely expressed as the minimum sum of products of the elements of $f(\vec{x})$ with [12]:

$$f(\vec{x})=\sum_{i=1}^{2^n-1}p_i m_i(\vec{x})\ ,\quad m_i=\prod_{j:b_j=1}x_j \tag{4}$$

where p_i is a Boolean variable deciding whether the *i*th Boolean product contributes to the result of applying function *f*. The *i*th product contains the input variables x_j for which the *j*th of the binary representation of the term's index i equals one.

In general, the number of PBFs (and therefore stack filters) is related to the number of input variables, namely the filter window size. Eq. (4) shows that an upper limit of PBFs of n input variables is 2^{2^n-1}, although a lower limit of $2^{2^n/2}$ can be derived by simplifying Eq. (4). Because of the computational complexity, it is much more difficult to design the large filters. In this paper, a new stack filter training algorithm is proposed. Each filter designed in this work is constructed by several 3×3 stack filters. This algorithm enables us to design stack filters with very large window sizes.

3 Proposed Algorithm for Designing Stack Filters with Large Window Sizes

3.1 Large Window Divided by Sub-windows

It is easy to find the optimal PBFs for the stack filters with small window sizes under the MAE criterion by the IMCSA. While the window sizes grow up to 5×5 or 7×7, it becomes very difficult to determine the optimal PBFs due to the sets of the stack filters that grow very rapidly. If the filters with large window sizes can be represented by these filters with smaller window size, such as 3×3, it may not be such difficult to configure. Our efforts were therefore directed toward the development of a representation of larger windows by smaller windows, which are called sub-windows in this paper. As an example, a 5×5 window can be constructed by several 3×3 windows, as is shown in Fig. 1.

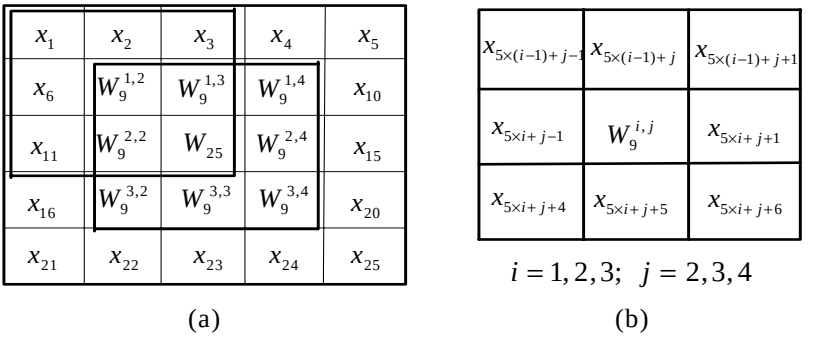


Fig. 1. (a) A 5×5 window divided by 3×3 windows. (b) A 3×3 sub-window.

As shown in Fig. 1, let $x_i, i=1,2,\cdots,25$ be the pixels of the 5×5 window, $W_9^{i,j}, i=1,2,3;\ j=2,3,4$ be the outputs of the filters with 3×3. The above statement can be formalized as:

$$W_9^{i,j}=S_{f_9}^r(x_{5\times(i-1)+j-1},\cdots,x_{5\times i+j+6}),\ i=1,2,3;\ j=2,3,4;r=1,2,\cdots,9, \tag{5}$$

where $S_f^r, r = 1, 2, \cdots, 9$ denote the corresponding stack filters with window size 3×3, the window as is shown in Fig. 1(b). In addition $W_9^{i,j}, i = 1, 2, 3;\ j = 2, 3, 4$ are also the pixels of the 3×3 window as is shown in Fig. 1 (a). W_{25} is the output of the 3×3 filter, which is called synthesis filter in this paper, formalized as Eq. (6).

$$W_{25} = S_{f_9}(W_9^{1,2}, W_9^{1,3}, \cdots, W_9^{3,4}), \tag{6}$$

where S_f is the 3×3 stack filter. Obviously W_{25} is determined by all the 25 pixels that contained in the 5×5 window, hence W_{25} is the output of the 5×5 filters. What we need to do is to design those 10 adaptive 3×3 stack filters. By dividing the 5×5 window into ten 3×3 sub-windows, the computational complexity of designing the 5×5 stack filters can be greatly reduced.

A 7×7 window can be divided by several 3×3 windows similarly, as is shown in Fig. 2.

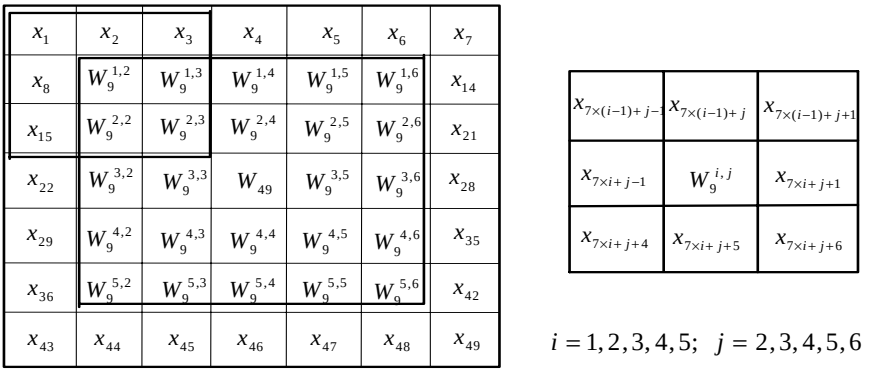


Fig. 2. (a) A 7×7 window divided by 3×3 windows. (b) The 3×3 windows.

where $x_i, i = 1, 2, \cdots, 49$ is the pixels of the original 7×7 window, $W_9^{i,j}, i = 1, 2, 3, 4, 5;\ j = 2, 3, 4, 5, 6$ is the output of the 3×3 stack filter. The corresponding window is shown in Fig. 2 (b). The statement can be formulized as Eq. (7).

$$W_9^{i,j} = S_f^r(x_{7\times(i-1)+j-1}, \cdots, x_{7\times i+j+6}),\ i = 1, 2, 3, 4, 5;\ j = 2, 3, 4, 5, 6; r = 1, 2, \cdots, 25, \tag{7}$$

W_{49} is the output of the 5×5 filter. The corresponding window is shown in Fig. 2(a), formalized as Eq. (8).

$$W_{49} = S_{f_{25}}(W_9^{1,2}, W_9^{1,3}, \cdots, W_9^{5,6}), \tag{8}$$

The steps for calculating the output of the 7×7 filters are as follows:

(1) Calculate the output of the 3×3 filters, $W_9^{i,j}, i=1,2,3,4,5;\ j=2,3,4,5,6$;

(2) Use the above algorithm to calculate the output of the filter with 5×5, namely the output of the filter with 7×7.

3.2 Stacking Property of the New Algorithm

In this section, we show that the filters produced by the proposed algorithm satisfy the stacking property.

Theorem 1: Assuming there are *m* positive Boolean functions $f_P^i(X_k^i), i=1,2,\cdots,m$ operating on threshold binary signals $X_k^i=(x_{1,k}^i, x_{2,k}^i,\cdots,x_{P,k}^i)$ in a length of *P*, where $x_{j,k}^i, j=1,2,\cdots,P,$ and $k=0,1,\cdots,K-1$ denote the binary values of the j^{th} sample at the k^{th} threshold level. If a Boolean function $f_N(X_k)$, operating on binary signals, $X_k=(x_{1,k}, x_{2,k},\cdots,x_{N,k})$, can be represented as, $f_N(X_k)=f_M(f_9^1(X_k^1), f_9^2(X_k^2),\cdots, f_9^m(X_k^m))$ where $X_k^i\in X_k, i=1,2,\cdots,m$, $f_M(\cdot)$ is a positive Boolean function with *m* input variables, then Boolean function $f_N(X_k)$ is also a positive Boolean function.

Proof: Let X_l be another threshold binary signals, $X_l=(x_{1,l}, x_{2,l},\cdots,x_{N,l})$ in a length of *N*, where $l=0,1,\cdots,K-1$, and $k<l$, $X_l^i=(x_{i_1,l}, x_{i_2,l},\cdots,x_{i_9,l})$, where $i_j=1,2,\cdots,N;\ j=1,2,\cdots,P$. The equality and inequality between them are defined as Eq. (9).

$$\begin{aligned} X_k=X_l &\Leftrightarrow x_{r,k}=x_{r,l} \text{ for } \forall r,\\ X_k\le X_l &\Leftrightarrow (x_{r,k}=1\Rightarrow x_{r,l} \text{ for } \forall r). \end{aligned} \tag{9}$$

Assume the binary sequence *X* possesses the stacking property, then $X_k\le X_l \Leftrightarrow k<l$. Then $X_k^i\le X_l^i, i=1,2,\cdots,m$. From the stacking property of the positive Boolean function, we obtain $f_P^i(X_k^i)\le f_P^i(X_l^i), i=1,2,\cdots,m$. Therefore the sequence constituted by the outputs of the positive Boolean functions $f_P^i(X_k^i), i=1,2,\cdots,m$ possesses the stacking property, as follows:

$$(f_P^1(X_k^1), f_P^2(X_k^2),\cdots, f_P^m(X_k^m)) \le (f_P^1(X_l^1), f_P^2(X_l^2),\cdots, f_P^m(X_l^m)). \tag{10}$$

And the Boolean function with *N* input variables possesses the stacking property as follows:

$$\begin{aligned} f_N(X_k)&= f_M(f_P^1(X_k^1), f_P^2(X_k^2),\cdots, f_P^m(X_k^m))\le\\ f_N(X_l)&= f_M(f_P^1(X_l^1), f_P^2(X_l^2),\cdots, f_P^m(X_l^m)) \end{aligned}. \tag{11}$$

Therefore, the Boolean function $f_N(X_k) = f_M(f_P^1(X_k^1), f_P^2(X_k^2), \cdots, f_P^m(X_k^m))$ is a positive Boolean function, and the filters produced by the proposed algorithm are thus stack filters.

3.3 Parallel Computing Features of the New Algorithm

In the above algorithm, those stack filters with smaller window sizes can be trained in parallel machines. Fig. 3 illustrates the parallel computing structure of the adaptive stack filtering algorithm.

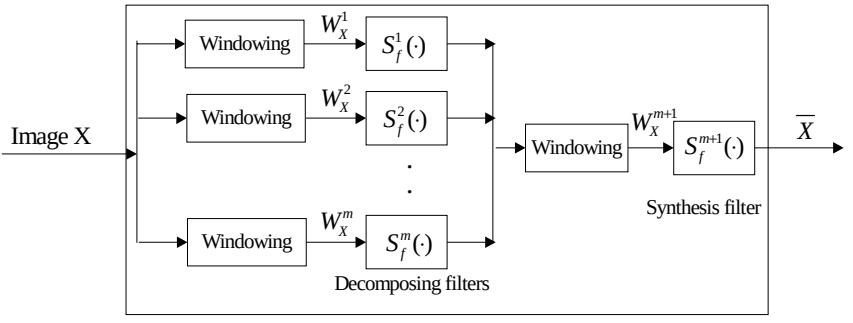


Fig. 3. Block diagram of optimal adaptive stack filtering algorithm

Fig. 3. shows that there are *m* decomposing filters for sub-windows, which can be trained independently. The PBF can be synthesized in a Look Up Table (LUT) which can be implemented in a Random Access Memory (RAM) [13]. Only 64 bytes are needed to implement a PBF of 9 input variables. For a PBF with 25 input variables, the number of RAM grows up to 4 MB. While the number of the PBF's input variables grow up to 49, more than 60,000 GB RAMs are required to implement the PBF. In this paper, the window sizes of the decomposing filters and synthesis filter are usually under 9-point, or 13-point. Thus the stack filtering algorithms with larger window sizes, such as 5×5, 7×7 can be easily implemented in FPGA using the proposed algorithm.

3.4 Applying IMCSA to Design 3×3 Stack Filters

By encoding a PBF into a binary string or some other forms, the task of finding an optimal PBF can be formulated as an optimization problem. And an improved Clonal Selection Algorithm [14], called Immune Memory Clonal Selection Algorithm (IMCSA) [11] is adopted to solve the optimization problem.
The main operations of IMCSA are:

(1) Generate a set (*A(0)*) of candidate solutions, select *s* solutions in *A(0)* as a memory cell (*M(0)*), set *k=0*;

(2) Select *t* best individuals of the populations (*A(k)* and *M(k)*), based on an affinity measure. Replace *t* members of population (*A(k)*) with the selected *t* individuals, resulting the refreshed antibody population (*A'(k)*);

(3) Reproduce (Clone) each individual of (*A'(k)*) , giving rise to a temporary population of clones (*Y(k)*). The clone size is an increasing function of the antibody-antigen affinity;

(4) Submit the population (*Y(k)*) to a genetic operation scheme with recombination, mutation. The mutation probability is proportional to the affinity of the antibody with the antigen. A maturated antibody population is generated (*Z(k)*);

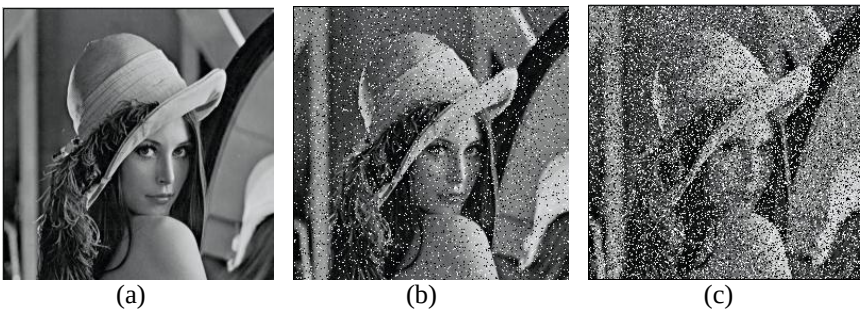
(5) Re-select the improved individuals from *Z(k)* to compose the new antibody population *A(k+1)*;

(6) Implement the immune operation on memory cell (*M(k)*), returning an improved memory cell (*M(k+1)*).

4 Experimental Results

In this section, experimental results are illustrated to demonstrate the validity and efficiency of the proposed algorithm to design the stack filters with large window sizes. Experiments were conducted on PC (P4 3.0 CPU) using C++ language programming.

In this experiment, the proposed algorithm is applied to a real image with size of 256×256 and 8-bit/pixel, which is corrupted with additive, uniformly distributed, uncorrelated impulse noise with the level from 10% to 60%. Our algorithms use a training window of $1/9^{th}$ full image area. Stack filters are trained using the original and noisy "Lena" images in Fig. 4(a), (b), and (c) and then applied to the noisy versions. The filtered images by the resulting stack filters are shown in Fig. 5. Their corresponding MAE's are shown in Table 1.



(a) (b) (c)

Fig. 4. Images used in experiments. (a) Original image "Lena" with 256×256. (b) "Lena" 10% noise. (c) "Lena"30% noise. The noisy images in (b) and (c) are corrupted by additive impulse noise with occurrence probability of 10% and 30% respectively.

Fig. 5(a), (b), and (c) show qualitative performance for 10% noise, and also show the effect of the size of stack filters, where for the 5×5 filter as well as 7×7 filter, the

images produced are visually more acceptable (fewer residual extreme values). Fig. 5(d), (e), and (f) are the resulting images by applying the stack filters to the image of Fig. 4(c). As expected, while the noise level grows up to 30%, the 7×7 stack filter has the best noise reduction performance. The experimental results show the good performance of our stack filtering algorithms. The relationship between MAE and noise probability is shown in Fig. 6, from which we find that the 7×7 stack filter has a great advantage over the 3×3 stack filter and 5×5 stack filter at high noise levels.

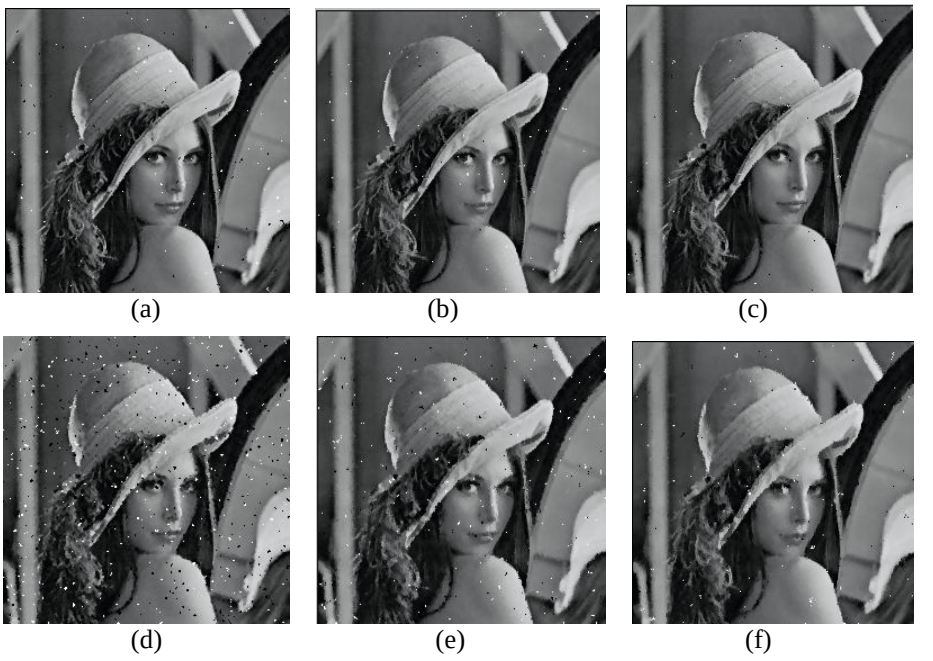


Fig. 5. (a) Output of a 3×3 stack filter designed by the new algorithm for "Lena" corrupted by 10% noise. (b) Same as (a) but with a 5×5 window. (c) Same as (a) but with a 7×7 window. (d) Output of a 3×3 stack filter designed by the new algorithm for "Lena" corrupted by 30% noise. (e) Same as (d) but with a 5×5 window. (f) Same as (d) but with a 7×7 window.

Table 1. Comparison in the performances of the MAE of stack filters with different window sizes

Corrupted Images	3×3 Median	5×5 Median	3×3 Stack filter	5×5 Stack filter	7×7 Stack filter
10%	3.8484	5.5615	2.29235	2.30637	2.53741
30%	8.2375	6.5647	6.91275	5.20232	5.02942

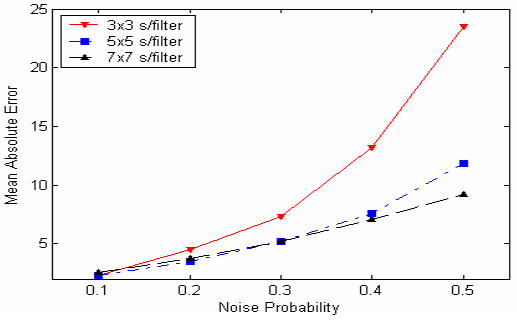


Fig. 6. MAE vari ation with stack filter sizes

5 Conclusions

In this paper, we proposed a new adaptive stack filtering algorithm for large window sizes. By dividing the large window into several sub-windows, the stack filter training process is significantly accelerated. And IMCSA is used to design the corresponding stack filters for the sub-windows. Then the desired stack filter is obtained by synthesizing these sub-filters. The new algorithm also preserves the inherent stacking property of stack filters.

Experiments with stack filters with 7×7 window are possible for the first time. The results show that the filters developed by this algorithm work well at high noise level.

References

1. P. D. Wendt, E. J. Coyle, and N. C. Gallagher, "Stack filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 898-911, Aug. 1986.
2. J. P. Fitch, E. J. Colyle, and N. C. Gallagher, "Median filtering by threshold decomposition," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 1183-1188, Dec. 1984.
3. L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo, "Weighted median filters: A tutorial," *IEEE Trans. Circuits Syst. II*, vol. 43, pp. 157-192, Mar. 1996.
4. J. W. Tukey, "Nonlinear (non-superposable) methods for smoothing data," *EASCON Conf. Rec.*, 1974.
5. T. Nodes and N. C. Gallagher, "Median filters: Some modifications and their properties," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 739-746, Oct. 1982.
6. P. A. Maragos and R. W. Schafer, "A unification of linear, median, order statistics, and morphological filters under mathematical morphology," *Proc, 1985 Int. Conf, Acoust., Speech, Signal Processing*, Tampa, FL. Mar. 1985.
7. G. Adams III, E. J. Coyle, L. Lin, L. Lucke, and K. Parhi, "Input compression and efficient VLSI architectures for rank order and stack filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 441-453, Aug. 1994.

8. J. Yoo, C. A. Bouman, E. J. Delp, and E. J. Coyle, "The nonlinear prefiltering and difference of estimates approaches to edge detection: Application of stack filters," *Graph. Models Image Process*., vol. 55, pp. 140-159, Mar. 1993.
9. I. Tabus, D. Petrescu, and M. Gabbouj, "A training framework for stack and Boolean filtering – Fast optimal design procedures and robustness case study," *IEEE Trans. Image Processing*, vol. 5, pp. 809 – 826, Jun. 1996.
10. J. Yoo, K. L. Fong, J. J. Huang, E. J. Coyle, and G. B. Adams, "A fast algorithm for designing stack filters," *IEEE Trans. Image Processing*, vol. 8, pp. 1014 – 1028, Aug. 1999.
11. W. S. Dong, G. M. Shi, and L. Zhang, "Immune memory clonal selection algorithm for designing stack filters," accepted by Neurocomputing.
12. S. Muroga, *Threshold logic and its applications*. New York: Wiley, 1971.
13. N. Woolfries et al, "Non Linear Image Processing on Field Programmable Gate Arrays," *NOBLESSE Workshop on Non-linear Model Based Image Analysis, Glasgow, Proc. NMBIA'98*, pp. 301-307, 1998.
14. L. N. De Castro and F. J. Von Zuben, "The clonal selection algorithm with engineering applications," *Proceedings of Genetic and Evolutionary Computation Conference, Workshop on Artificial Immune Systems and Their Applications*, pp. 36-37, Jul. 2000.

Fast Segmentation of Cervical Cells by Using Spectral Imaging Analysis Techniques

Libo Zeng[1] and Qiongshui Wu[2]

[1] College of Computer Science, Wuhan University, Wuhan, 430072, P.R. China
lbzeng@whu.edu.cn
[2] College of Electronic Information, Wuhan University, Wuhan, 430079, P.R. China
qswu@whu.edu.cn

Abstract. Cervical cancer is the second most common cancer among women worldwide. Early detection of cervical cancer is very important for successful treatment and increasing survival. We report a spectral imaging microscopic system for Papanicolaou smear analysis for early detection of cervical cancer. Different from traditional color imaging method, we use spectral imaging techniques for image acquisition, which can simultaneously record spectral and spatial information of a sample. In this paper, the imaging instrument construction and spectral image acquisition method is introduced. In the image segmentation process, an effective algorithm using spectral ratio method is applied for cell nuclei detection, which can easily detect the nuclei and diminish the influence of the cytoplasm overlap. Results showed that our segmentation is robust and precise. In addition to this, the segmentation speed is very high.

Keywords: Cervical cell, spectral imaging, spectral image segmentation.

1 Introduction

1.1 Early Detection of Cervical Cancer

Currently, *Papanicolaou (Pap)* test is the most popular and effective screening test for cervical cancer[1]. In many Western countries, cervical cancer screening programs (based on cytology-Pap smears) have reduced cervical cancer incidence and mortality by as much as 90 percent[2].

With recent advance in technology, computer-assisted apparatus for automated Pap screening can release human from burdensome repetitive labor and increase detection rates of abnormalities. In the most of current automated systems, image acquisition and segmentation is the most important. Since most of the cells are overlapped with each other in Pap smear images, it is challenging to detect nucleus precisely. In many systems, they are analyzed as a group or directly discarded, which will decrease the sensitivity of the whole system. If these overlapped cells are precisely segmented, there will be a significant improvement.

Due to these matters, we proposed a new image analysis system for early cervical cancer detection based on Pap smears Different from traditional color imaging method, we use spectral imaging.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 734 – 741, 2006.

1.2 Spectral Imaging

Spectral imaging is a novel imaging technology that can simultaneously record spectral and spatial information of a sample[3]. A spectral image consists of a serial of grey images, each acquired at a desired narrow band of wavelengths. As shown in Fig. 1(a), the spectral image can be looked as a 3-dimensional image cube(x-y-λ dimensions), the 2-dimensional image records the spatial information of a sample, and the third dimension represents spectral wavelength. Each pixel is associated with a spectrum. As shown in Fig. 1(b), by plotting the grey intensity of the images in the cube at the same position sequently versus wavelength, the spectrum at this point will be obtained. Hence, a spectral image can show not only the spatial features of the target, but also can reveal the unique spectral information of each pixel.

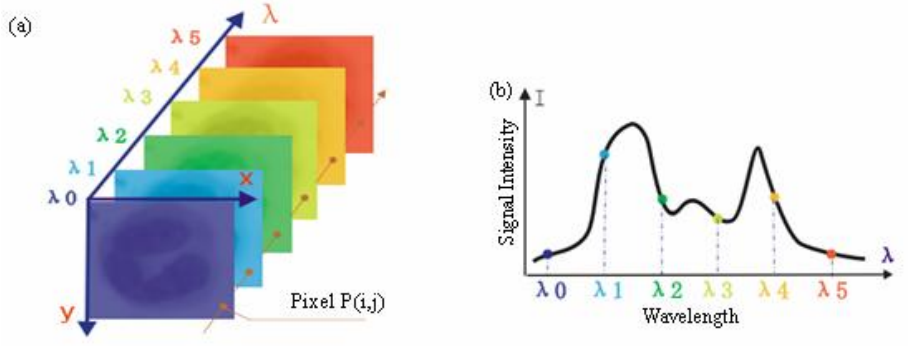


Fig. 1. Principle of spectral imaging. (a) Spectral image cube. (b) Spectrum at pixel P(i,j).

The combination of spectroscopy and image analysis technology can take advantage of current various tools borrowed from spatial image processing, chemometrics and specific spectroscopy. As we all know that biological tissue exhibits unique spectra in transmission. Spectral differences in tissue pathology may be spatially resolved using imaging processing and spectral analysis techniques. The additional spectral information can make our segmentation method more flexible. What's more, for traditional color imaging method with R, G and B channels, a lot of chemical and physical information is buried under the spectral characteristics of the imaging system. Using spectral imaging can remedy these problems, and thus increase the color image quality significantly.

2 Instrument

A schematic diagram of the spectral imaging apparatus is shown in Fig. 2.

This apparatus consists of a BX41 microscope (Olympus Corporation), a *Liquid Crystal Tunable Filter (LCTF)* device and its controller (VariSpecTM, Cambridge Research &Instrumentation, Inc.), a three-dimensional automation stage, a Pentium IV

computer (PC), a cooled monochrome *Charge Coupled Device (CCD)* camera (1024*1024, 12bits, Penguin 600L, Pixera Corporation) and etc.

A 6V/30W halogen lamp is used as a bright-field source with Koehler illumination, which is connected to a stabilized power supply. The objective of 40x /0.65 NA is used and the condenser aperture iris diaphragm is set to 75% of the NA.

A LCTF device[4] is placed in the imaging path, one side coupled with a CCD Camera and the other side mounted on the microscope by the standard C-mount. The image signal of the sample is filtered by the LCTF.

A set of image processing and analysis software package was developed using Microsoft Visual C++ 6.0. It not only covers general image processing and analysis functions, but also has functions of spectral analysis, including spectral image acquisition, illumination and system response calibration, color image reconstruction, and spectral analysis.

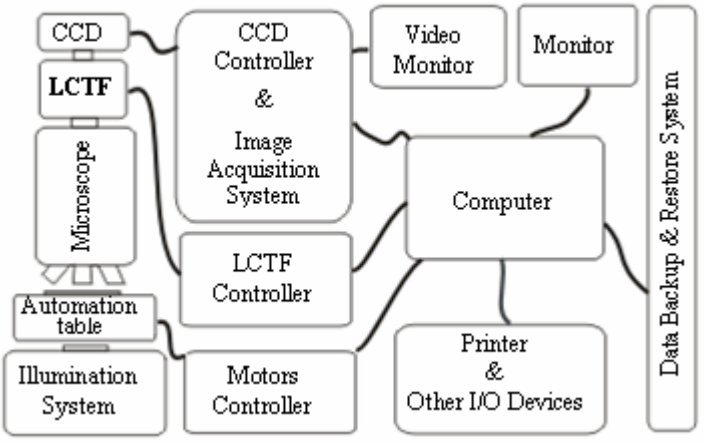


Fig. 2. Schematic diagram of a spectral imaging microscope apparatus based on LCTF

3 Multispectral Image Acquisition

The LCTF is used to provide wavelength selection from range 400 nm to 700 nm with an increment of 10 nm, an image cube is acquired by sequently commanding the filter to a desired wavelength, taking an image, and this process repeats until all the images under desired wavelengths are captured. All the images of the same field of view were stacked in a computer, from the lowest wavelength to the highest, to create an “image cube”, which is called spectral image.

4 Segmentation

For segmentation, Pap smear images are segmented into cervical cells nucleus, cytoplasm, and background regions. Since it is difficult to detect nucleus and cytoplasm

directly, taking this into consideration, we do it in an indirect way: Background regions are firstly deleted, and of course the left regions S_1 are the cell nucleus and cytoplasm regions. And then, binary operation is performed for each patch of cell clusters individually to separate nucleus and cytoplasm.

Different from traditional color or grey image segmentation methods, which are so excessively dependent on absolute grey intensity of image, our segmentation is mainly based on spectral ratio image produced by a division operation between images at different wavelengths.

4.1 Background Segmentation

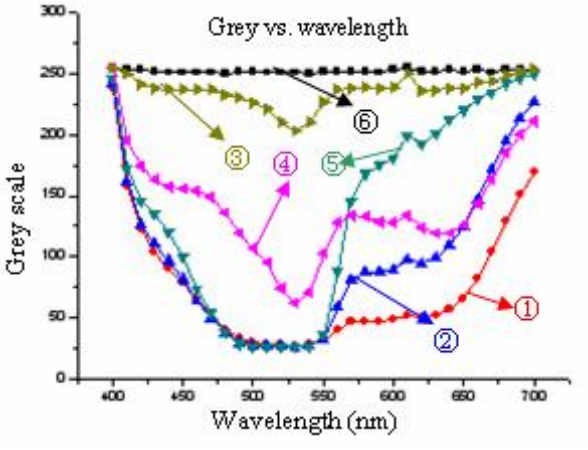


Fig. 3. Typical spectrum of cervical cells in Pap smear images. ①②nucleus ③④⑤cytoplasm ⑥background.

Observing the spectrum of cervical cells in spectral Pap smear image, as shown in fig. 3, we can find that different components of a cell have different spectrum. Nucleus has the lowest grey intensity, and the background has the highest grey intensity.

To find background, from Fig. 3, we can find that at 530nm the different between background and other components is the most remarkable, thus, just apply an absolute threshold to image of 530nm can bring us good preliminary segmentation results. Experiments showed that this method is very effective for images of good quality, but not very robust when the smears are not well preprocessed or there is impurity. Taken this into consideration, we sought for better segmentation method. Experiments show that better results can be obtained by using spectral operation method.

With spectral ratio method, a new gray image $G(i, j)$ can be generated:

$$G(i, j) = \min\left(\max\left(\rho \frac{W(i, j, \lambda_m) - W(i, j, \lambda_n)}{W(i, j, \lambda_n)}, 0\right), 255\right) \quad (1)$$

$W(i, j, \lambda_m)$ represents the gray value of a pixel*(i,j)* in the image of wavelength λ_m, ρ is the factor for setting the new image limited to a given maximal value of 255. With

different wavelengths λ_m and λ_n carefully selected, our interested components in the image can be highly enhanced.

In our segmentation, ρ is set to the value of 1800 according to masses of experiments, λ_m is 580nm and λ_n is 530nm.

After the ratio image is generated according to equation (1), a medial filter is applied for smoothing (Result is shown in fig. 4(b)), and then, an automatic thresholding method of Otsu[5] is used to delete the background regions, thus, a mask regions S_1 of nucleus and cytoplasm regions is formed.

From fig. 4(b), we can find that the nucleus and cytoplasm regions are greatly enhanced.

4.2 Nucleus Enhancement

Since the cervical cytoplasm and nucleus have a big variety of color, common image segmentation methods which only use color information are almost impossible to differentiate them. From fig. 4(a), we can easily find that the colors of nucleus vary in a large range. Some are very dark while some are in light dark, and some are buried under the overlapped regions of cytoplasm. For cytoplasm, its color covers the range from light blue to thick red, and there are a lot of regions have thick color than nucleus. For above reasons, nucleus detection is a very important subject in cervical smear analysis.

Different from the most of common color

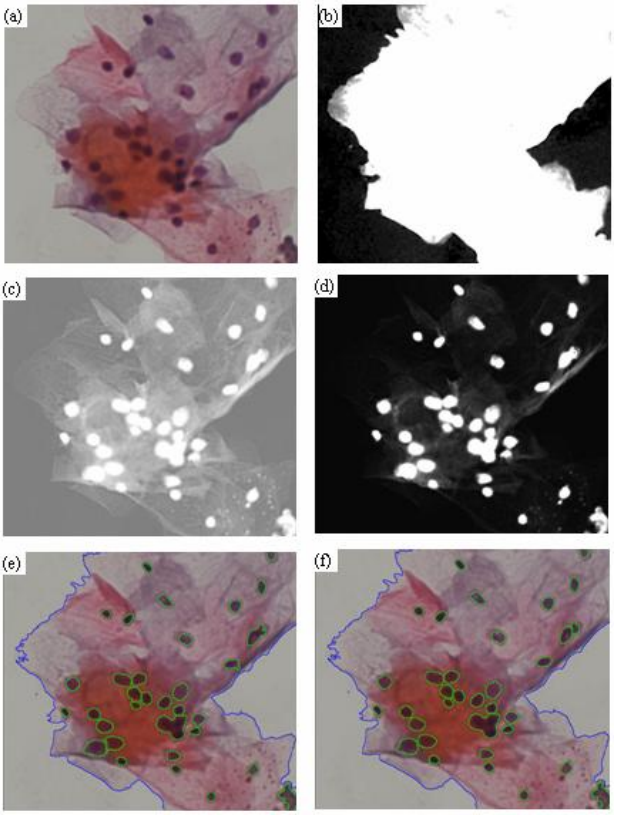


Fig. 4. Cervical cells segmentation process. (a) Original Color image fused by multispectral image cube. (b) background enhancement. (c)image of 700nm/600nm. (d) Gamma adjustment for image (c). (e)Preliminary segmentation result (f) Final results.

image segmentation methods that use HSV (Hue, Saturation, Value) spaces converted from RGB (Red, Green, Blue) spaces, we use spectral ratio image analysis method. The ratio image is generated as equation (3) shows:

$$G(i, j) = \min\left(\rho \frac{W(i, j, \lambda_m)}{W(i, j, \lambda_n)}, 255\right) \tag{2}$$

In our experiments, ρ is set to 128, λ_m and λ_n is set to 700nm and 600nm individually.

Fig. 4(c) is the ratio image generated as equation (2). From fig. 4(c), we can find that the nucleus regions are well enhanced. Both the black nucleus and the fawn nucleus almost have the same grey intensity and all the nucleus regions are outstanding. Even the nuclei (Marked with arrows) that are buried under the overlapped cytoplasmare also enhanced.

In order to further enhance it, a Gamma adjustment process is followed for nucleus enhancement (Result is shown in fig. 4(d)). From fig. 4(d), we can find that the nucleus regions are enhanced remarkably and the grey intensity of cytoplasm are cut down greatly.

4.3 Nucleus Segmentation

The nucleus segmentation process is as following:

(1) Delete background. In this step, a mask region S_1 contains nucleus and cytoplasm regions is formed.

(2) Nucleus enhancement as described in chapter 4.2.

(3) For each patch in region S_1, using an automatic thresholding method of Otsu to find a threshold to segment nucleus out. In this step, the nucleus mask S_2 is formed.

(4) A watershed algorithm is implemented to S_2 for declustering of the inevitable clusters. And then delete small particles, and then mask S_3 is formed. In this step, satisfactory preliminary results can be obtained [As fig. 4(e) shows]. But in many cases, the boundary of many nuclei is not very precise. So we need further precisely segmentation.

(5) For each patch in regions S_3, firstly find the circumscribed rectangle, and then enlarge the rectangle with 4 pixels on each side. And then apply an automatic thresholding method of Otsu in the new rectangle for binarization. In this step, the new nucleus mask S_4 is formed.

(6) Apply an open binary morphological operation to S_4, and then using watershed algorithm to declustering of the inevitable clusters, and then delete small particles. In this step, the final results (As fig. 4(f) shows) are obtained.

From fig. 4(f), we can find that the final segmentation results are highly satisfactory. The nuclei in deep dark or carmine, or even buried under the overlapped regions of cytoplasm can be well segmented out.

4.4 A Comparison with Traditional Color Analysis Methods

Most of reported work for cervical cancer detection is based on traditional color image analysis methods, there are very few reports is based on spectral image.

Since most of color image analysis method is based on HSV model. Here, we will have a comparison with traditional HSV model. Compared with HSV model, our method have many advantages: (1) Firstly, we use spectral ratio images for analysis while traditional HSV analysis uses original Hue, Saturation or Value channel for analysis. That is to say, the targets for our analysis are relative values but not the original absolute grey values. (2) Our images are spectral images. Each spectral image contains a set of grey images of different wavelengths. A spectral image can produce many ratio images when two different wavelengths are selected. With special different wavelengths carefully selected, our interested objects can be enhanced dramatically. From fig. 4(b)(c), we can easily find that the cytoplasm and nucleus are well enhanced as our desire, which is hardly obtained by HSV analysis method. For such an enhancement extent, there are no better reports found before. (3) Compared with the HSV images, our spectral ratio image is more direct and convenient for further analysis, and more similar to original image in shape when enhanced.

5 Conclusions and Future Work

Our spectral ratio segmentation method can effectively diminish the influences of illumination inconsistencies, CCD's quantum efficiency (QE) curves, optic ununiformity and throughput properties, which makes our images segmentation device-independent in a large extent. For traditional image analysis method, even the same computational programme may yield quite different results when the imaging conditions change. These problems can be well overcome with our method.

A mass of experiments show that our segmentation results are highly satisfactory. It is more robust and precise than conventional color imaging method while with high speed. It proved that our spectral imaging analysis introduced into the early detection of cervical is successful.

Compared to traditional imaging analysis method, our segmentation has following advantages: 1) Segmentation is more flexible and accurate. Spectral images consist of both spectral information and image information. All common image processing techniques can be applied to them, in additional to this, current powerful spectral analysis techniques can also be applied to them. The additional spectral information is helpful for precise segmentation. 2) The segmentation method is more robust and insensitive to staining and imaging environments, which makes the segmentation results easily reproducible, for our spectral images are device-independent in a broad imaging environment and our segmentation is based on the ratio image, not the absolute gray value of original images. 3) With low computational cost. Our segmentation algorithm is very straightforward, by using a simple spectral division and automatic thresholding operation, we can readily obtain satisfactory preliminary segmentation results. With our segmentation method, any image of 1024*1024 pixels can be segmented in 0.3 to 3 seconds with our Pentium IV1.7GHz CPU.

Though preliminary results show that our spectral imaging analysis method in early detection of cervical cancer is promising and inspiring, yet for a practical system, there are still a lot of room for improvements: (1) Wavelengths selection. Currently, the wavelengths are selected manually according to large numbers of experiments, we believe that there will be an automatic selection method, which can save us a lot of time and provide us with more optimal results. (2) Further boost segmentation speed. Currently the speed is high comparing to previous reported method, but for a real practical system, it is not enough. Our future solution is to implement the segmentation operation by using special-designed hardware, such as DSP card. (3) Perform a study on 100x magnification images analysis. The advantage is that it can increase the scan speed than current 400x. A preliminary experiment shows that it is feasible and practical. 4) More effective feature extraction is one of the endeavors in future developments.

References

[1] R. A. Smith, C. J. Mettlin, K. J. Davis, etc., "American Cancer Society guidelines for the early detection of cancer", CA Cancer J Clin, **50**:34–49, 2000.

[2] L. Peter Ph.D., M.Sc, "A System for Automated Screening for Cervical Cancer", Visible Diagnostics A/S, http://www.imm.dtu.dk/visiondag/VD03/medicinsk/pl.pdf, June 11, 2003.

[3] R. M. Levenson, C. C. Hoyt, "Spectral imaging and microscopy", AMERICAN LABORATORY, November, pp. 26-33, 2000.

[4] Cambridge Research & Instrumentation, CRI Varispec Tunable Filter User's Manual, http://WWW.CRI-INC.COM

[5] N. Otsu, "A threshold selection method from gray level histograms", *IEEETrans System Man Cybernetics SMC*, **8**: 62-66, 1978.

Local Geometry Driven Image Magnification and Applications to Super-Resolution

Wenze Shao[1] and Zhihui Wei[2]

[1] Department of Computer Science and Engineering, Nanjing University of Science and Technology, 210094 Nanjing, China
shaowenze@qianlong.com
[2] Graduate School, Nanjing University of Science and Technology, 210094 Nanjing, China
gswei@mail.njust.edu.cn

Abstract. Though there have been proposed many magnification works in literatures, magnification in this paper is approached as reconstructing the geometric structures of the original high-resolution image. The structure tensor is able to estimate the orientation of both the edges and flow-like textures, which hence is much appropriate to magnification. Firstly, an edge-enhancing PDE and a corner-growing PDE are respectively proposed based on the structure tensor. Then, the two PDE's are combined into a novel one, which not only enables to enhance the edges and flow-like textures, but also to preserve the corner structures. Finally, the novel PDE is applied to image magnification. The method is simple, fast and robust to both the noise and the blocking-artifact. Another novelty in the paper is the application of the novel PDE to super-resolution reconstruction, plus additional term for image fidelity. Experiment results demonstrate the effectiveness of our approach.

1 Introduction

Image processing has been one of the hottest research topics currently. Large numbers of image processing approaches are emerged [1-4]. Image magnification is just such an interesting problem which mainly aims at producing the original high-resolution (HR) image from a single low-resolution (LR) and perhaps noisy image. Taking into account the insufficient density of the imaging sensor, it is reasonable to consider the observation model: $\vec{\boldsymbol{g}} = \boldsymbol{D}\vec{\boldsymbol{u}} + \vec{\boldsymbol{n}}$, where $\vec{\boldsymbol{g}}$, $\vec{\boldsymbol{u}}$, and $\vec{\boldsymbol{n}}$ are respectively the column-ordered vectors of the $M \times N$ LR image g, the original $qM \times qN$ HR image u, and the additive random noise n. Besides, the variable q represents the undersampling factor, and the matrix $\boldsymbol{D}$ describes the nonideal sampling process, i.e., first local-average and then down-sampling.

There have been proposed many magnification algorithms [5-10] in literatures, which formulate image magnification as an ill-posed inverse problem. While currently the PDE-based level-set approaches [11-12] are the most popular choices, which are fast, edge-enhancing and robust to the noise. Generally, the level-set magnification approaches can be unified to the following PDE

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 742 – 751, 2006.

$$\partial u / \partial t = c_1 D^2 u(\eta,\eta) + c_2 D^2 u(\xi,\xi) .$$

with the initial image as the bilinear or bicubic interpolation of the LR image. In the above PDE, $\eta = Du / |Du|$, $\xi = Du^{\perp} / |Du|$ are orthonormal vectors in the direction of gradient and tangent respectively, and c_1, c_2 are the diffusivity functions of the gradient $|\nabla u|$ in each direction. Nevertheless, as for the case of nonideal sampling, image sharpening has to be incorporated into the interpolation process. Therefore, we propose the following PDE through incorporating in the shock filtering [13-14]

$$\partial u / \partial t = c_1 D^2 u(\eta,\eta) + c_2 D^2 u(\xi,\xi) - \beta sign(D^2 u(\eta,\eta)) |\nabla u| . \quad (1)$$

where β is a positive controlling parameter. PDE (1) essentially magnifies images driven by the level curves, however, level curves do not capture all the geometric information that one desires to analyze the image content. Hence, more flexible PDE's should be exploited to tackle with different geometric structures.

Image magnification in this paper is approached as reconstructing the geometric structures of the original HR image. The structure tensor is able to estimate the orientation of both the edges and flow-like textures, which hence is much appropriate to magnification. Firstly, an edge-enhancing PDE and a corner-growing PDE are respectively proposed based on the structure tensor. Then, the two PDE's are combined into a novel one, which not only enables to enhance the edges and flow-like textures, but also to preserve the corner structures. Finally, the novel PDE is applied to image magnification. The method is simple, fast and robust to both the noise and the blocking-artifact. Another novelty in the paper is the application of the novel PDE to super-resolution reconstruction, plus additional term for image fidelity. Experiment results demonstrate the effectiveness of our approach.

The paper is organized as follows. Section 1 gives a brief review of the previous PDE-based magnification algorithms. In section 2, an edge-enhancing PDE and a corner-growing PDE are respectively proposed based on the structure tensor, which are then combined into a novel one. Subsequently, section 3 applies the novel PDE to image magnification and section 4 applies it to super-resolution image reconstruction. Concluding remarks are finally given in section 5.

2 Edge-Enhancing PDE and Corner-Growing PDE

To estimate the orientation of local geometric structures, Weickert [15] proposed the well-known structure tensor

$$J_\rho(\nabla u_\sigma) = G_\rho * (\nabla u_\sigma \otimes \nabla u_\sigma), \ \rho \geq 0 . \quad (2)$$

denoted as $(J_{m,n})_{m=1,2;n=1,2}$, where u_σ is the regularized version of u with a Gaussian kernel $N(0,\sigma^2)$, making the edge detection insensitive to the noise at scales smaller than σ; the tensor $\nabla u_\sigma \otimes \nabla u_\sigma$ is convolved by a Gaussian kernel $N(0,\rho^2)$, making the structure analysis more robust to the flow-like structures and noises. The matrix J_ρ is symmetric and definite semi-positive, and hence has orthonormal eigenvectors, denoted as w and $w^{\perp}$ respectively. The vector w, defined as

$$w = \begin{pmatrix} 2J_{12} \\ J_{22} - J_{11} + \sqrt{(J_{22} - J_{11})^2 + 4J_{12}^2} \end{pmatrix}, \quad w = w / |w|. \tag{3}$$

points in the direction with the largest contrast, and the orthonormal vector $w^{\perp}$ points in the structure direction. Their corresponding eigenvalues μ and $\mu^{\perp}$ can be used as descriptors of local structures: Constant areas are characterized by $\mu = \mu^{\perp} = 0$, straight edges by $\mu \gg \mu^{\perp} = 0$, and corners by $\mu \geq \mu^{\perp} > 0$. Besides, $(\mu - \mu^{\perp})^2$ is the measure of the local coherence, represented by $(\mu - \mu^{\perp})^2 = (J_{22} - J_{11})^2 + 4J_{12}^2$.

2.1 Edge-Enhancing PDE

Based on the two eigenvectors w and $w^{\perp}$ provided by the structure tensor J_{ρ}, we generalize PDE (1) to the following PDE

$$\partial u / \partial t = c_1 D^2 u(w, w) + c_2 D^2 u(w^{\perp}, w^{\perp}) - \beta sign(D^2 u(w, w)) |\nabla u|. \tag{4}$$

where β is a positive parameter, c_1 and c_2 are respectively the diffusivity functions of the local coherence $(\mu - \mu^{\perp})^2$. Particularly, when c_1 is a monotonically decreasing function ranged from 1 to 0, and c_2 is equal to 1, the behavior of PDE (4) is easy to interpret. On homogeneous regions of u , the function c_1 has values near 1. Then the first two terms of the PDE will be combined into Δu , yielding the isotropic diffusion. As for regions with edges and flow-like textures, the function c_1 has values near 0 and hence the first term of PDE (4) vanishes. The second term forces PDE (4) to smooth the image in the structure direction, and therefore preserves the edges and flow-like textures. The last term in PDE (4) corresponds to the shock filtering for image sharpening, while the vector η in PDE (1) is replaced by the vector w . In fact, we notice that the same modification has also been proposed in [16]. Though PDE (4) may perform well on preserving the edges and flow-like textures, while how about the corner structures. Fig. 1(d) shows that PDE (4) is not capable of preserving the corner structures.

2.2 Corner-Growing PDE

To overcome the blurring effect of corner structures in image diffusion, we propose the following PDE for corner growing

$$\partial u / \partial t = c_3 \ (\nabla u)^{\mathrm{T}} \cdot (\nabla \cdot (w^{\perp} \otimes w^{\perp})). \tag{5}$$

where c_3 is a positive controlling parameter, and the divergence operator ∇ for a 2×2 matrix M is defined as

$$\nabla \cdot M = \begin{pmatrix} \nabla \cdot (m_{11} & m_{12})^{\mathrm{T}} \\ \nabla \cdot (m_{21} & m_{22})^{\mathrm{T}} \end{pmatrix}, \quad M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}.$$

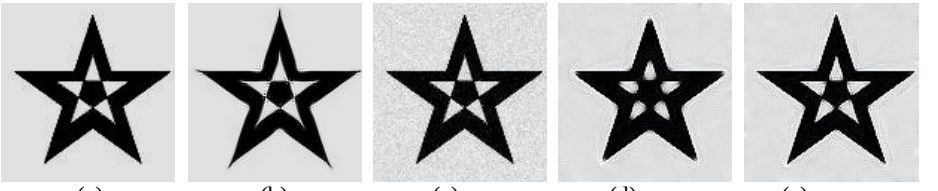


Fig. 1. (a) Original image, (b) Diffuse Fig.1 (a) with PDE (5), (c) Gaussian noisy image (μ = 0, σ = 10), (d) Diffuse Fig.1 (c) with PDE (4), (e) Diffuse Fig.1 (c) with PDE (6)

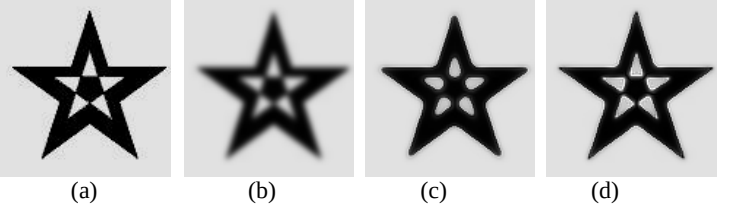


Fig. 2. (a) Original image, (b) Convolved image by the Gaussian kernel (σ = 2), (c) Diffuse Fig.2 (b) with PDE (7), (d) Diffuse Fig.2 (b) with PDE (6)

Here, we demonstrate the performance of PDE (5) utilizing an interesting experiment as shown in Fig. 1. Obviously, Fig. 1(b) shows that PDE (5) plays the role of corner growing, and in fact, the rate of corner growing is determined by both parameters ρ and c_3. Therefore, we combine PDE's (4) and (5), obtaining a novel PDE (6) which not only enables to enhance the edges and flow-like textures, but also to preserve the corner structures

$$\partial u/\partial t = \begin{cases} c_1 D^2 u(w,w) + c_2 D^2 u(w^\perp, w^\perp) + c_3 \ (\nabla u)^{\mathrm{T}} \cdot (\nabla \cdot (w^\perp \otimes w^\perp)) \\ -\beta sign(D^2 u(w,w)) \,|\nabla u| \end{cases}. \tag{6}$$

Other than PDE (6), Weickert [15] proposed the following PDE

$$\partial u/\partial t = \nabla \cdot (D(J_\rho(\nabla u_\sigma))\nabla u) .$$

Where D is defined as $D(J_\rho(\nabla u_\sigma)) = c_1(w \otimes w) + c_2(w^\perp \otimes w^\perp)$, called diffusion tensor. To achieve image sharpening, the modified shock filtering can be also incorporated in the above PDE, obtaining

$$\partial u/\partial t = \nabla \cdot (D(J_\rho(\nabla u_\sigma))\nabla u) - \beta sign(D^2 u(w,w)) \,|\nabla u| . \tag{7}$$

Nevertheless, PDE (6) is more powerful than PDE (7) in preserving corner structures. Experiment results shown in Fig. 2 tell us the truth.

3 Local Geometry Driven Image Magnification

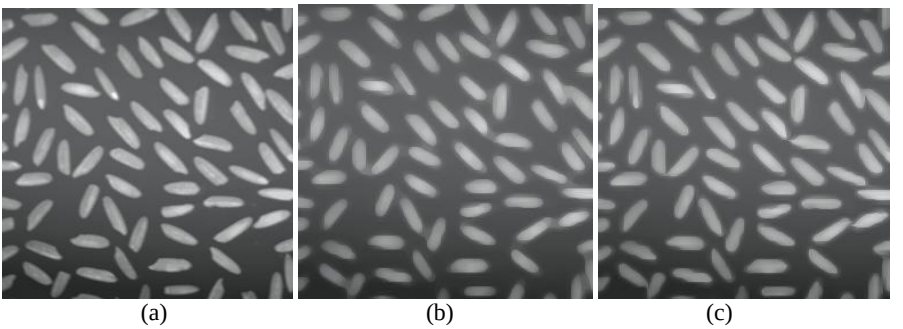
In this section, we make use of PDE (6) for image magnification, with $u(x,0)=u_0(x)$ as the initial image (bilinear or bicubic interpolation of the LR image g). By now, each term in PDE (6) has had its corresponding physical meaning in magnification: the first and second term combine to perform the isotropic diffusion in the homogeneous regions; the second term plays the role of smoothing the blocking artifacts in the structure direction; the third term plays the role of preserving the corner structures; while the fourth term overcomes the blurring effect introduced in the interpolation process. Since PDE (6) considers almost all of the geometric structures in images, which hence is much more appropriate to magnification than the level-set approaches.

For a vector $\varpi=(\varpi_1,\varpi_2)^{\mathrm{T}}$, $D^2u(\varpi,\varpi)=(\varpi_1)^2u_{xx}+2\varpi_1\varpi_2u_{xy}+(\varpi_2)^2u_{yy}$. And as for the first-order partial derivatives ∂_x and ∂_y, they are calculated by recently proposed optimized derivative filters [17], with properties of rotation invariance, accuracy and avoidance of blurring effects. The corresponding numerical scheme of PDE (6) is as follows:

$$u_x^{t+1}=u_x^t+\tau\left\{\begin{matrix}c_1D^2(u_\sigma^t)_x(w,w)+c_2D^2(u_\sigma^t)_x(w^\perp,w^\perp)+\ c_3(\nabla u_x^t)^{\mathrm{T}}\cdot(\nabla\cdot(w^\perp\otimes w^\perp))\\+\beta(-sign(D^2(u_\sigma^t)_x(w,w)))\left|\nabla u_x^t\right|\end{matrix}\right\}. \tag{8}$$

where τ is the size of each iteration step, x is the pixel location and t is the current iteration time. Then, PDE (6) can be implemented by the following steps:

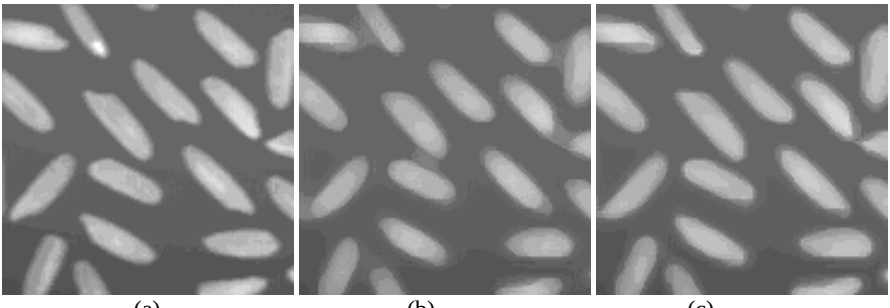
1. Calculate the initial image $u_0(x)$ using bilinear interpolation;
2. Calculate the structure tensor $J_\rho(\nabla u_\sigma)=G_\rho*(\nabla u_\sigma\otimes\nabla u_\sigma)$ using (2);
3. Calculate the dominate vector w using (3);
4. Calculate the local coherence and diffusivity functions c_1 and c_2;
5. Calculate $(u_\sigma)_x$, $(u_\sigma)_y$, $(u_\sigma)_{xx}$, $(u_\sigma)_{xy}$, $(u_\sigma)_{yy}$, $D^2u_\sigma(w,w)$, and $D^2u_\sigma(w^\perp,w^\perp)$;
6. Calculate u_x, u_y, $|\nabla u|$ and $(\nabla u)^{\mathrm{T}}\cdot(\nabla\cdot(w^\perp\otimes w^\perp))$;
7. Update the iteration process using (8) (The number of iteration step is T).



(a) (b) (c)

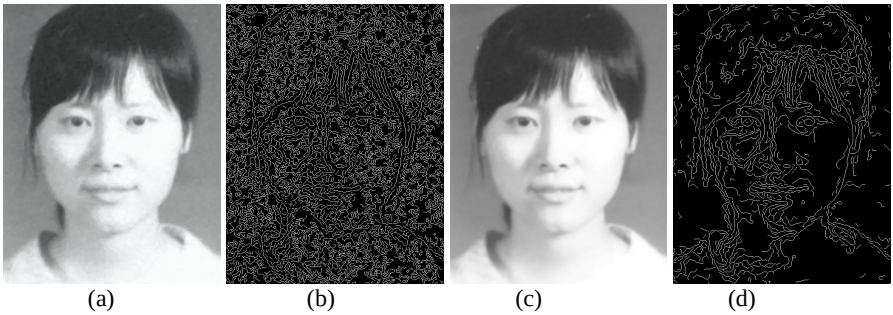
Fig. 3. (a) Original image, (b) Level-set approach (T = 20, $\tau=0.24, c=1, c_2=1, \beta=0.15$), (c) Our proposed approach (T = 20, $\tau=0.24, \sigma=1.5, \rho=2, c=1, c_2=1, c_3=1.5, \beta=0.15$)

The diffusivity function c_1 in PDE's (1) and (6) is chosen as $c_1(t) = c/(1+t^2)$ for gray images ranged from 0 to 255, and c_2 is defined as a variable. Hence, there are overall 8 parameters in the numerical scheme (8): T , τ , σ , ρ , c , c_2 , c_3 , β . For simplification, three parameters are fixed in the following experiments, including: $\tau = 0.24$, $\rho = 2$, $\sigma = 1.5$.



(a) (b) (c)

Fig. 4. Magnified portions corresponding to (a) Fig. 3(a), (b) Fig. 3(b), (c) Fig. 3(c)



(a) (b) (c) (d)

Fig. 5. (a) Bicubic interpolation, (b) Edge detection of Fig. 5(a) using Canny operator, (c) Our proposed interpolation ($T = 20, \tau = 0.24, \sigma = 1.5, \rho = 2, c_2 = 1, c_3 = 1, \beta = 0.15$), (d) Edge detection of Fig. 5(c) using Canny operator

The shock filtering-incorporated level-set approach (1) and our proposed PDE (6) are utilized for image magnification. The initial guess for both PDE (1) and PDE (6) is chosen as the bilinear interpolation of the LR image. Fig. 3 shows the magnification results when the undersampling factor is 4. And Fig. 4 shows the portions of Fig. 3. Obviously, our proposed approach achieves much better visual quality compared with the level-set approach. As a matter of fact, the level-set approach not only removes the corners, but also shortens the level curves of the original HR image. Fig. 5 shows the magnification results of a real-scanned photo as the undersampling factor is 3, utilizing the bicubic interpolation and our proposed approach respectively. It is clear that the novel approach also behaves more robust in suppressing the random noises, as well as enhancing the edges.

4 Applications to Super-Resolution Image Reconstruction

Super-resolution image reconstruction aims at producing one or a set of high-resolution images from a set of geometric warped, blurred, noisy, and under-sampled low-resolution images. In the last two decades, there has been proposed a variety of super-resolution methods [18-21]. Almost all of them originate in the theory of regularization since super-resolution reconstruction is mathematically an ill-posed inverse problem. Thus, the novel PDE can also be applied to super-resolution.

Given a sequence of degraded LR frames $\{g_k\}_{k=1,\ldots,p}$, the imaging process of g_k from the super-restored image u can be formulated as follows

$$\vec{\boldsymbol{g}}_k = \boldsymbol{D}_k \boldsymbol{B}_k \boldsymbol{M}_k \vec{\boldsymbol{u}} + \vec{\boldsymbol{n}}_k \quad \Rightarrow \quad \vec{\boldsymbol{g}}_k = \boldsymbol{H}_k \vec{\boldsymbol{u}} + \vec{\boldsymbol{n}}_k \,. \tag{9}$$

where $\vec{\boldsymbol{u}}$ is the vector column-reordered from the HR image u of size $[N \times N]$, $\vec{\boldsymbol{g}}_k$ is the vector column-reordered from the LR image g_k of size $[L \times L]$ ($L < N$), $\vec{\boldsymbol{n}}_k$ is the column-ordered additive Gaussian noise of size $[L \times L]$, $\boldsymbol{M}_k$ is the warp matrix of size $[N^2 \times N^2]$, $\boldsymbol{B}_k$ is the blur matrix of size $[N^2 \times N^2]$, and $\boldsymbol{D}_k$ is the decimation matrix of size $[L^2 \times N^2]$. Degradation model (9) can be further written as

$$\begin{bmatrix} \vec{\boldsymbol{g}}_1 \\ \vdots \\ \vec{\boldsymbol{g}}_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{D}_1 \boldsymbol{B}_1 \boldsymbol{M}_1 \\ \vdots \\ \boldsymbol{D}_p \boldsymbol{B}_p \boldsymbol{M}_p \end{bmatrix} \vec{\boldsymbol{u}} + \begin{bmatrix} \vec{\boldsymbol{n}}_1 \\ \vdots \\ \vec{\boldsymbol{n}}_p \end{bmatrix} \quad \Rightarrow \quad \vec{\boldsymbol{g}} = \boldsymbol{H} \vec{\boldsymbol{u}} + \vec{\boldsymbol{n}} \,. \tag{10}$$

In literatures [18-22], edge-preserving super-resolution is often approached as the following minimization energy functional

$$\vec{\boldsymbol{u}} = arg \min_{\vec{\boldsymbol{u}}} \left\{ \frac{\lambda}{2} \cdot \left\| \vec{\boldsymbol{g}} - \boldsymbol{H} \vec{\boldsymbol{u}} \right\|^2 + \boldsymbol{E}_{Regularization}[\vec{\boldsymbol{u}}] \right\} . \tag{11}$$

where $\boldsymbol{E}_{Regularization}[\vec{\boldsymbol{u}}]$ is the regularization term for edge-preservation. In general, steepest descent is exploited to solve the minimization functional, and then super-resolution reconstruction is approximately equivalent to

$$u_x^{t+1} = u_x^t + \tau \left\{ \lambda \cdot \boldsymbol{H}^T \left(\vec{\boldsymbol{g}} - \boldsymbol{H} \vec{\boldsymbol{u}}^t \right) \big|_x + \left(-\nabla E_2[u^t] \big|_x \right) \right\} . \tag{12}$$

where τ is a scalar determining the step size in the gradient direction, t is the iteration time, and x is the pixel position in the HR image plane. The matrix-vector calculation in (12) is implemented as the direct image operators such as warp, blur, and decimation [23], which is an extremely fast and memory efficient way.

Unlike the minimization functional (11), we approach the super-resolution reconstruction as the following formulation, based on our proposed PDE (6)

$$\frac{\partial u}{\partial t} = \begin{Bmatrix} c_1 D^2 u(w,w) + c_2 D^2 u(w^\perp, w^\perp) + c_3 \ (\nabla u)^{\mathrm{T}} \cdot \left(\nabla \cdot \left(w^\perp \otimes w^\perp \right) \right) \\ -\beta sign\left(D^2 u(w,w) \right) |\nabla u| + \lambda \boldsymbol{H}^T \left(\vec{\boldsymbol{g}} - \boldsymbol{H} \vec{\boldsymbol{u}}^t \right)(x) \end{Bmatrix} . \tag{13}$$

The corresponding numerical scheme of PDE (13) is the same as (8), except the last term for image fidelity. According to the discussion on PDE (6) in section 2, it is

easy to conclude that the proposed approach is not only capable of edge-preservation, but also edge-enhancement, corner-preservation and image sharpening. Hence, PDE (13) is more appropriate to the super-resolution task, especially for the cases of large sampling factors and noise levels. The simulation result demonstrates the strength of the proposed algorithm (All the experiments are implemented by MATLAB 7.0 on Pentium(R) 4 CPU 2.4 GHz).

Fig. 6. Super-resolution results (a) Original image (256×256), (b) One of LR frames (64×64; Noise variance is 5), (c) the TV approach (PSNR=28.0098), (d) Our proposed approach (PSNR=28.6500)

In the experiment, we create a sequence of LR frames utilizing an HR image [Fig. 6(a)]. First, we shifted the HR image by a pixel in the vertical direction. Then, the shifted image was locally averaged and under-sampled by the factor of 4 in each direction. The same operation with different motion vectors in vertical and horizontal

directions was used to produce 8 LR frames from the original scene. Finally, the resulting LR frames were added to the white Gaussian noise (The noise variance is 5). One of these LR frames is presented in Fig. 6(b) (pixel replication). Fig. 6(c) shows the super-resolved HR image obtained by the TV approach [22] and the steepest descent (12). Fig. 6(d) is the result obtained by our proposed approach. In the TV approach, the experiment parameters are: T=15, τ=0.24, and λ=0.15; and in the proposed method, the experiment parameters are: T=15, τ=0.24, σ=1.5, ρ=2, c=5, c_2=1, c_3=1, β=0.05, and λ=0.25. To objectively evaluate the quality of images, the metric of PSNR (peak signal-to-noise ratio) [24] is used. As for Fig. 6(c), PSNR=28.0098; and as for Fig. 6 (d), PSNR=28.6500. Hence, the proposed approach is more powerful in keeping the image fidelity, and also suppressing well the random noises. On the other hand, compared with Fig. 6(c), it is obvious to see that Fig. 6(d) has higher contrast, smoother edges, and more natural appearance. Therefore, the proposed approach achieves better visual quality than the TV approach.

5 Conclusions

The paper proposed an alternative PDE approach for image magnification based on the proposed edge-enhancing PDE and corner-growing PDE, which is not only capable of enhancing the edge structures, but also preserving the corner structures. The method is simple, fast and robust to both the noise and the blocking-artifact. Another novelty in the paper is the application of the novel PDE to super-resolution image reconstruction, plus additional term for image fidelity. Experiment results demonstrate the effectiveness of our approach.

References

1. Tan, Y.P., Yap, K.H., Wang, L.P. (Eds.): Intelligent Multimedia Processing with Soft Computing. Springer, Berlin Heidelberg, New York (2004)
2. Vernazza, G. (Eds.): The IEEE International Conference on Image Processing. Genoa Italy (2005)
3. Huang, D.E. (Eds.): The International Conference on Intelligent Computing. Springer, Berlin Heidelberg, Hefei China (2005)
4. Campilho, A.; Kamel, M.: Image Analysis and Recognition. Springer, Berlin Heidelberg, Portugal (2006)
5. Blu, T., Th´evenaz, P., Unser, M.: Linear Interpolation Revisited. IEEE Transactions on Image Processing, Vol. 13 (2004) 710–719
6. Li, X., Orchard, T.: New Edge-Directed Interpolation. IEEE Transactions on Image Processing, Vol. 10 (2001) 1521–1527
7. El-Khamy, S.E., Hadhoud, M.M., Dessouky, M.I., Salam, B.M., El-Samie, F.E.: Efficient Implementation of Image Interpolation as an Inverse Problem. Digital Signal Processing, Vol. 15 (2005) 137–152
8. Schultz, R.R., Stevenson, R.L.: A Bayesian Approach to Image Expansion for Improved Definition. IEEE Transactions on Image Processing, Vol. 3 (1994) 233–242
9. Guichard F., Malgouyres F.: Total Variation based Interpolation. EUSIPSO, III (1998) 1741–1744

10. Chan, T.F., Shen, J.H: Mathematical Models for Local Nontexture Inpaintings. SIAM J. Appl. Math., Vol. 62, No. 3 (2002) 1019–1043
11. Belahmidi, A., Guichard F.: A Partial Differential Equation Approach to Image Zoom. Proceedings of International Conference on Image Processing (2004)
12. Morse, B.S., Schwartzwald, D.: Isophote-based Interpolation. In 5th IEEE International Conference on Image Processing (1998)
13. Osher, S.J., Rudin, L.I.: Feature-Oriented Image Enhancement Using Shock Filters. SIAM J. Numer. Anal., Vol. 27 (1990) 919–940
14. Alvarez, L., Mazorra, L.: Signal and Image Restoration Using Shock Filters and Anisotropic Diffusion. SIAM J. Numer. Anal., Vol. 31, No. 2 (1994) 590–605
15. Weickert, J.: Coherence-Enhancing Diffusion Filtering. International Journal of Computer Vision, Vol. 31, No. 2/3 (1999) 111–127
16. Weickert, J.: Coherence-Enhancing Shock Filters. Pattern Recognition, Lecture Notes in Computer Science, Vol. 2781. Springer-Verlag, Berlin Heidelberg (2003) 1–8
17. Weickert, J.: A Scheme for Coherence-Enhancing Diffusion Filtering with Optimized Rotation Invariance. Journal of Visual Communication and Image Representation, Vol. 13, No. 1/2 (2002) 103–118
18. Borman, S., Stevenson, R.L.: Super-resolution for Image Sequences—A Review. In Proc. IEEE Int. Symp. Circuits and Systems (1998) 374–378
19. Farsiu, S., Robinson, M.D.: Fast and Robust Multiframe Super Resolution. IEEE Transactions on Image Processing, Vol. 13, No.10 (2004) 1327–1344
20. Park, S.C., Park, M.K., Kang M.G.: Super-resolution Image Reconstruction – A Technical Overview. IEEE Signal Processing Magazine, Vol. 20, No. 3 (2003) 21–36
21. Nguen, M.K., Bose, N.K.: Mathematical Analysis of Superresolution Methodology. IEEE Signal Processing Magazine (2003) 62–74
22. Capel, D., Zisserman, A.: Super-resolution Enhancement of Text Image Sequences. Proceedings of International Conference on Pattern Recognition (2000) 600–605
23. Zomet, A., Peleg, S.: Efficient Super-resolution and Applications to Mosaics. In Proc. Int. Conf. Pattern Recognition (2003) 579-583
24. Banham, M.R., Katsaggelos, A.K.: Digital Image Restoration. IEEE Trans. Signal Processing (1997) 24-41

Three Dimensional Image Inpainting

Satoru Morita

Faculty of Engineering, Yamaguchi University
2-16-1 Tokiwadai Ube, 755-8611, Japan
smorita@yamaguchi-u.ac.jp

Abstract. Recently the method restoring an old picture using the local differential equation on the basis of the geometric measure is proposed. It is necessary to restore an old film and a medical image with noise as well as an old picture. So we extend the method applying for the two dimentional image such as a picture to the three dimensional image such as a time sequence image and a medical image. It is necessary to obtain an object boundary from the original image in order to generate a composite image of good quality, which is difficult to distinguish from the original image. If an object boundary can not be detected, it is difficult to remove the object. In this study, we propose a method for detecting an object boundary and removing it and inpainting its image in a manner that makes it difficult to distinguish from the original image. We extend the image partition method based on the level set method to the method applying for the movie and the medical image to detect an object boundary. We demonstrate its effectiveness by removing a terop from a movie and a tumor from a three dimensional medical image.

1 Introduction

Inpainting technique can be utilize for restorting an old picture[1]. Recently, the inpainting method using the local differential equation based on the geometric measure is proposed[2][3][4]. The neighbourhood pixels of an object boundary in the two dimensional image is used for the image inpainting method.

It is necessary to detect an object boundary to generate the good composite image such that the person is not conscious that it is composite image. We cannot generate the composite image if we cannot detect an object boundary. So it is important to automatically detect an object bounary from the noisy image. The active contour model is proposed to detect an object from the noisy image[5]. This method is extend to the method applying for the three dimensional image[6]. While the region detection using the moving curved surface based on the local differntial equation is proposed[7]. The method of the region partition using the local differential equation based on the geometric measure with phases is applied for two dimensional image[8][9] but is not not applied for three dimensional image. So we extend the method applying in the two dimensional image to the three dimensional image for detecting an object boundary in the three dimensional image.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 752–761, 2006.

We apply for both the time sequence image and the medical image which are three dimensional image.

The method using space feature [2][4][10][11] and the method using time feature [12][13] are studied. In this method, we regard the time sequence image as three dimensional image. We detect an object boundary and inpaint the object using both time and space features. If an object appears and disappears, the object is inpainted using the adjacents image. If else, an object is inpainted using the neibourhood of the object boundary in the current image. Thus, the time and space features are efficiently used for the time sequence image.

The example of the medical image inpainting does not exist while some researchers study to automatically detect a focus from the medical image. The method based on moving curved surface using the differential equation is studied [14][15]. That is why an object boundary can be detected from the noisy image. We extend the image partition method based on the geometric measure applied for the two dimensional image to the three dimensional image. The medical image which a focus is removed can be appeared before a surgical operation. The technique can be utilize the virtual surgery simulation[16][17][18].

Though sequential image in a video is improved by using image at neighbor time in the video inpainting method [19][20]reported recently, the method detecting an object boundary is not discussed and the method can not be applied in the three dimensional density image such as CT and MRI image in which three axis is common in the property. On the other hand , we do not pay attention to the difference between time and image though our method is applied for the video inpainting. If we take into consideration with it, our method can be improved for video inpainting.

In section 2 , we explain the three dimensional region detection, the three dimensional inpainting in section 3 and we show that the effectiveness of this method by removing the terop from the movie in section 4.

2 Three Dimensional Image Segmentation

Firstly, we explain the method for detecting an object boundary in a two dimensional image[8][9]. Recently, an image processing method based on the moving curved surface according to the partial differential equation was proposed. This method is referred to as the level set method because we use the partial differential equation defining the level of $\phi = 0$. For this study, we use an active contour model based on the partial differential equation by defining the level of $\phi = 0$. The contour in the image generated by deforming a curved surface is defined according to the partial differential equation. The original image is u_0, and the phase ϕ is the function required for the region detection and can define the initial value arbitrarily. The boundary of the object is the phase $\phi = 0$ in the image. u is the image produced as a result of region partition.

$$\frac{\partial \phi}{\partial t} = \delta(\phi)(\nu \nabla \cdot (\frac{\nabla \phi}{|\nabla \phi|}) - ((u_0 - c_1)^2 - (u_0 - c_0)^2)) \tag{1}$$

$$c_1 = \frac{\int u_0 H(\phi) dxdy}{\int H(\phi) dxdy}$$

$$c_0 = \frac{\int u_0 (1 - H(\phi)) dxdy}{\int (1 - H(\phi)) dxdy}$$

$$u = c_1 H(\phi) + c_0 (1 - H(\phi)) \tag{2}$$

This method can be extend for detecting a number of regions. The function H is referred to as the Heviside function, and it is satisfies $\delta(x) = \frac{\partial H(x)}{\partial x}$. This holds true for the method is different from the active contour model based on the geometric measure, the method for useing some phases ϕ to detect some regions. An image can be divided into the regions of 2^n regions by using n number of phases ϕ. The image is divided into 2^n regions. The method of using the partial differential equation to divide the image into four regions is shown below.

$$\frac{\partial \phi_1}{\partial t} = \delta(\phi_1)(\nu \nabla \cdot (\frac{\nabla \phi_1}{|\nabla \phi_1|})$$

$$- (((u_0 - c_{11})^2 - (u_0 - c_{01})^2) H(\phi_2) + ((u_0 - c_{10})^2 - (u_0 - c_{00})^2)(1 - H(\phi_2)) \tag{3}$$

$$\frac{\partial \phi_2}{\partial t} = \delta(\phi_2)(\nu \nabla \cdot (\frac{\nabla \phi_2}{|\nabla \phi_2|})$$

$$- (((u_0 - c_{11})^2 - (u_0 - c_{10})^2) H(\phi_1) + ((u_0 - c_{01})^2 - (u_0 - c_{00})^2)(1 - H(\phi_1)) \tag{4}$$

$$c_{11} = \frac{\int u_0 H(\phi_1) H(\phi_2) dxdy}{\int H(\phi_1) H(\phi_2) dxdy}$$

$$c_{10} = \frac{\int u_0 H(\phi_1)(1 - H(\phi_2)) dxdy}{\int H(\phi_1)(1 - H(\phi_2)) dxdy}$$

$$c_{01} = \frac{\int u_0 (1 - H(\phi_1)) H(\phi_2) dxdy}{\int (1 - H(\phi_1)) H(\phi_2)) dxdy}$$

$$c_{00} = \frac{\int u_0 (1 - H(\phi_1))(1 - H(\phi_2)) dxy}{\int (1 - H(\phi_1))(1 - H(\phi_2)) dxdy} \tag{5}$$

$$u = c_{11} H(\phi_1) H(\phi_2) + c_{10} H(\phi 1)(1 - H(\phi_2))$$

$$+ c_{01}(1 - H(\phi_1)) H(\phi_2) + c_{00}(1 - H(\phi_1))(1 - H(\phi_2)) \tag{6}$$

Next, we extend the method for detecting the regions to a three dimensional image. Since two phases are used in this experiment, the image is classified into four regions. If three phases are used, the image will be classified into eight regions. We arbitrarily generate the objects arbitrarily for each phase. The ϕ sign is positive in the interior of an object and the sign of ϕ is negative outside the object. The initial phase ϕ of a pixel is defined as the distance between the $\phi = 0$ contour and the pixel.

$$c_{11} = \frac{\int u_0 H(\phi_1) H(\phi_2) dxdydz}{\int H(\phi_1) H(\phi_2) dxdydz}$$

$$c_{10} = \frac{\int u_0 H(\phi_1)(1 - H(\phi_2)) dxdydz}{\int H(\phi_1)(1 - H(\phi_2)) dxdydz}$$

$$c_{01} = \frac{\int u_0 (1 - H(\phi_1)) H(\phi_2) dxdydz}{\int (1 - H(\phi_1)) H(\phi_2)) dxdydz}$$

$$c_{00} = \frac{\int u_0 (1 - H(\phi_1))(1 - H(\phi_2)) dxdydz}{\int (1 - H(\phi_1))(1 - H(\phi_2)) dxdydz} \tag{7}$$

$$u = c_{11} H(\phi_1) H(\phi_2) + c_{10} H(\phi 1)(1 - H(\phi_2))$$
$$+ c_{01}(1 - H(\phi_1)) H(\phi_2) + c_{00}(1 - H(\phi_1))(1 - H(\phi_2)) \tag{8}$$

3 Three Dimensional Image Inpainting

The inpainting method is proposed for the two dimensional image. The u image is the original image that required the inpainting.

$$\frac{\partial u}{\partial t} = \nabla^{\perp} \cdot \nabla \Delta u + \nu \nabla \cdot (g(|\nabla u|) \nabla u) \tag{9}$$

Fig. 1. The image sequences in a movie with terop

The symbol $\nabla^{\perp}$ represents $(-\frac{\partial}{\partial y}, \frac{\partial}{\partial x})$. The image is inpainted according to the partial differential equation.

We show an algorithm based on the partial differential equation. The gray value is represented as $u(i, j)$ in the (i, j) pixel. This image $u(i, j)$ changes as time t increases, according to the following equation.

$$\frac{\partial u}{\partial t} = u_t^n(i, j) \tag{10}$$

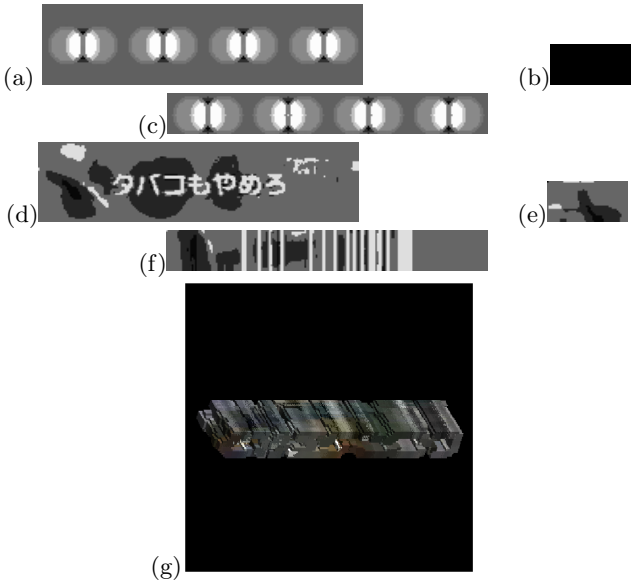


Fig. 2. Figures (a), (b), and (c) show the x-y, y-z,and z-x cross sections of the initial regions detected from the movie with the terop at the iteration number $t = 0$. Figures (d), (e), and (f) show the x-y, y-z, and z-x cross sections of the regions detected from the movie terop at the iteration number $t = 4096$. (g) The detected three dimensional boundary of the terop.

where the situation of the surface with equal gray level in the (i, j) pixel is defined as $L^n(i, j)$. This is shown using the following equation.

$$L^n(i,j) = u^n_{xx}(i,j) + u^n_{yy}(i,j) \tag{11}$$

where the changing value δL^n of the diffused gray level is obtained using L^n.

$$\delta L^n = (L^n(i+1,j) - L^n(i-1,j), L^n(i,j+1) - L^n(i,j-1)) \tag{12}$$

We calculate the difference between the gradient direction of the gray level and the direction that we would like to diffuse.

$$\beta^n(i,j) = \delta L^n(i,j) \cdot \frac{N(i,j,n)}{|N(i,j,n)|} \tag{13}$$

$$\frac{N(i,j,n)}{|N(i,j,n)|} = \frac{(u^n_x(i,j), u^n_y(i,j))}{\sqrt{((u^n_x(i,j))^2 + (u^n_y(i,j)^2))}} \tag{14}$$

$$u^n_t(i,j) = \beta^n(i,j) \cdot |\nabla u^n(i,j)| \tag{15}$$

The image is inpainted by repeating this algorithm.

Fig. 3. The image sequences removed terop from the movie

The algorithm is extended to apply to the three dimensional image $u(i, j, k)$. The density is given as the $u(i, j, k)$ image in the (i, j, k) pixel. The image changes according to the following equation.

$$\frac{\partial u}{\partial t} = u_t^n(i, j, k) \tag{16}$$

The situation of the surface with equal gray level at the (i, j, k) pixel is defined as $L^n(i, j, k)$ This is shown using the following equation.

$$L^n(i, j, k) = u_{xx}^n(i, j, k) + u_{yy}^n(i, j, k) + u_{zz}^n(i, j, k) \tag{17}$$

The changing value δL^n of the diffused gray level is obtained using L^n.

$$\delta L^n = (L^n(i+1, j, k) - L^n(i-1, j, k), L^n(i, j+1, k) - L^n(i, j-1, k),$$
$$L^n(i, j, k+1) - L^n(i, j, k-1)) \tag{18}$$

We calculate the difference between the gradient direction of the gray level and the direction that we would like to diffuse.

$$\beta^n(i, j, k) = \delta L^n(i, j, k) \cdot \frac{N(i, j, k, n)}{|N(i, j, k, n)|} \tag{19}$$

$$\frac{N(i, j, k, n)}{|N(i, j, k, n)|} = \frac{(u_x^n(i, j, k), u_y^n(i, j, k), u_z^n(i, j, k))}{\sqrt{((u_x^n(i, j, k))^2 + (u_y^n(i, j, k)^2) + (u_z^n(i, j, k)^2))}} \tag{20}$$

$$u_t^n(i, j, k) = \beta^n(i, j, k) \cdot |\nabla u^n(i, j, k)| \tag{21}$$

The inpainting of the three dimensional image proceeds by iterating this calculation. The object is removed from the three dimensional image, and we paint the interior of the object using the background in a discrete manner.

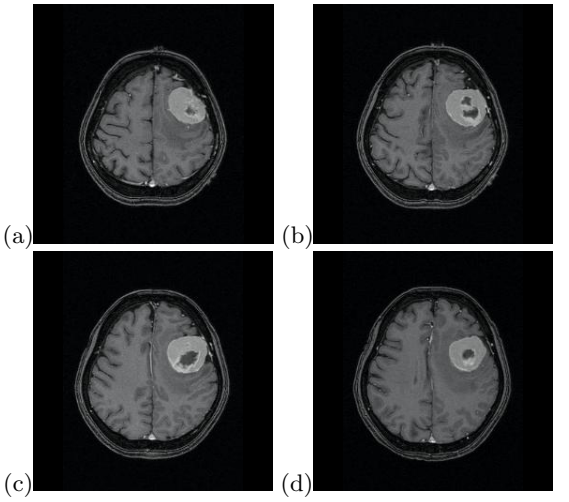


Fig. 4. Figures (a), (b), (c), and (d) show the cross section of the original medical MRI image showing a brain tumor appears for $z = 30$, $z = 33$, $z = 36$, and $z = 39$ respectively

4 Generating Three Dimensional Composite Image

Next, we explain the process flow for removing an object from a three dimensional image.

- An object boundary is detected from the image by using the region partition method based on the partial differential equation, which in turn is based on a geometric measure. (2.)
- An object is inpainted using the image inpainting method. (3.)

Figure 1 shows five images from a movie. The image size is 640×400 pixels, and the three dimensional image is composed of 1000 images. The size of the three dimensional image is $640 \times 400 \times 100$ pixels. We show five images from among the thousand images. The parameter $\nu = 0.03$ and the time constant value $d\tau = 0.0001$ are used. Figure 2(a), (b), and (c) show the x-y, y-z, and z-x cross sections for the initial regions detected from the movie with the terop at the iteration number $t = 0$. Figures 2 (d), (e) and (f) show the x-y, y-z, and z-x cross sections for the regions detected from the movie with the terop at the iteration number $t = 4096$. Figure 2 (g) shows the detected three dimensional boundary of the terop. Figure 3 shows the five images without the terop is removed that corresponding to the five images of figure 1 in the movie using three dimensional image inpainting. It is found that you cannot find the terop in the result movie and the terop is inpainted using the background. The black edges exist in the boundary between the terop and the background in the movie using the experiment to emphasize the terop. The black edge is detected using the expansion process of the morphological method before the three dimensional

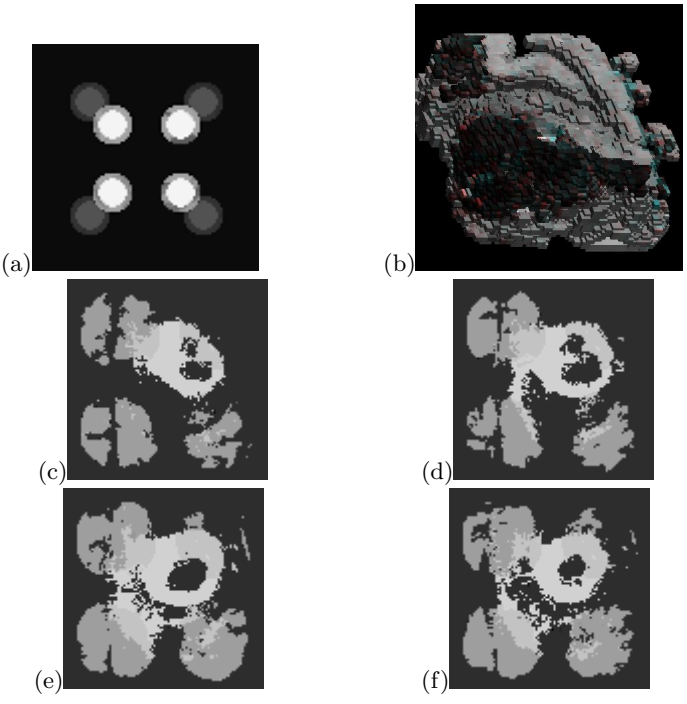


Fig. 5. The detected brain tumor. (a) The cross section of the initial regions for $z = 35$ at the iteration number $t = 0$ of medical MRI image in which a brain tumor appears. (b) The detected three dimensional boundary of the brain tumor. Figures (c), (d), (e), and (f) show the cross section for the detected regions for $z = 30$,z=33, $z = 36$, and $z = 39$ at the iteration number $t = 4096$ respectively.

image inpainting. It is found that the movie without the terop is generated from the original movie with terop. The terop existed in three dimensional image is detected from the sequential images using three dimensional method and the terop is inpainted and romoved using two dimensional method.

Figure 4 shows a three dimensional medical image of a brain tumor. Figure 4(a), (b), (c), and (d) are the cross section of the initial regions for $z = 30, z = 33, z = 36$, and $z = 39$ for a original three dimensional medical image with a brain tumor respectively. The size of three dimensional image is $100 \times 100 \times 60$ pixels. Figure 5 shows the medical image of the brain tumor. Figure 5(a) shows the cross section of the initial regions for $z = 35$ at the iteration number $t = 0$. Figure 5(b) is the detected three dimensional boundary of the brain tumor. Figure 5(c), (d), (e), and (f) show the cross sections of the detected regions for $z = 30$, $z = 33$, $z = 36$ and $z = 39$ at the iteration number $t = 4096$. It is found that the brain tumor is detected. The parameter $\nu = 0.03$ and the time constant $d\tau = 0.0001$ are used. Figure 6 shows the composite image generated from figure 5. Figure 6(a), (b), (c), and (d) show the cross sections of the medical MRI image without brain tumor which is removed for $z = 30$, $z = 33$, $z = 36$,

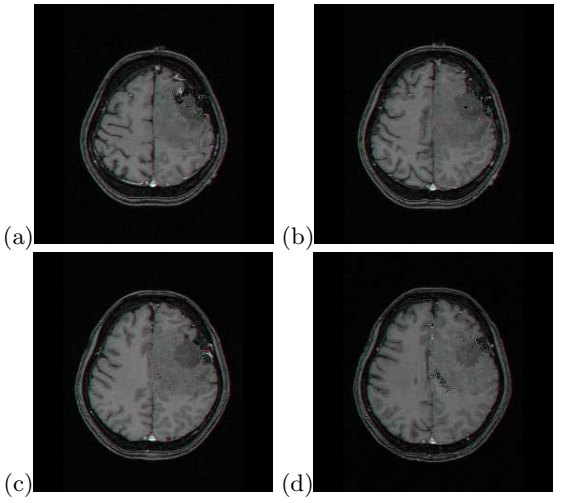


Fig. 6. Figures (a), (b), (c), and (d) show the cross section of the medical MRI image without the brain tumor which is removed using three dimensional image inpaiting for $z = 30$, $z = 33$, $z = 36$, and $z = 39$ respectively

and $z = 39$ respectively. It is found that the white focus shown in figure 4 is romoved in figure 6. A focus is inpainted by detecting its boundary in the three dimensional image and using the background of this image.

5 Conclusions

We showed that a three dimensional composite image can be automatically generated by detecting an object boundary and removing the object before the inpainting.

References

1. Kokaram, A.C., Morris, R.D., Fitzgerald, W.J., Rayner, P.J.W.: Detection of missing data in image sequences. IEEE Transactions on Image Processing **11** (1995) 1946–1508
2. Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. SIGGRAPH2000 (2000) 417–424
3. Bertalmio, M., Bertozzi, A.L., Saphiro, G.: Navier-stokes, fluid dynamics and image and video inpainting. CVPR 2001 (2001) 353–363
4. Dori, I., Cohen-Or, I., Yeshurun, H.: Fragment-based image completion. SIGGRAPH2003 (2003) 303–312
5. Kass, M., A, W., Terzopoulos, D.: Snakes: Active contours models. International Journal of Computer Vision **1** (1988) 321–322
6. Cohoen, L.D.: On active contour models and balloons. CVGIP: Image Understanding **53** (1991) 211–218

7. Sethian, J.: Curvature and the evalution of fronts. Comm. Math. Physics **101** (1985) 487–499
8. Vese, L.A., Chan, T.F.: A multiphase level set framework for image segmentation using the mumford and stah model. International Journal of Computer Vision **50** (2002) 271–293
9. Mumford, D., Stah, J.: Optimal approximation by piecewise smooth function adn associated variational problems. Comm. Pure Appl. Math. **42** (1989) 577–685
10. Efros, A., Leung, K.: Texture synthesis by non-parametric sampling. Proc. of International Conference on Computer Vision **2** (1999) 1033–1038
11. Criminisi, A., Perez, P., Toyama, K.: Object removal by exemplar-based inpainting. Proc. of IEEE Computer Vision and Pattern Recognition **2** (2003) 721–728
12. Wang, J., Adelson, E.: Ieee transactions on image processing special issue: Image sequence compression. CVGIP: Image Understanding **3** (1994) 625–638
13. Irani, M., Rousso, B., Peleg, S.: Computing occluding and transparent motions. International Journal of Computer Vision **12** (1994) 5–16
14. Malladi, R., Sethian, J.A., Vemuri, B.C.: Shape modeling with front propagation: A level set approach. IEEE Trans. of Pattern Analysis and Machine Intelligence **17** (1995) 158– 175
15. Han, X., Prince, J.L.: A topology preserving level set method for geometric deformable models. CVGIP: Image Understanding **25** (2003) 755–768
16. Bro-Nielsen, M.: Finite element modeling in surgery simulation. Proc. IEEE **86** (1993) 490–503
17. Delingette, H.: Toward realistic soft-tissue odeling in medical simulation. Proc. IEEE **86** (1998) 512–523
18. Gross, M.H.: Computing occluding and transparent motions. Computer graphics in medicine: From visualization to surgery simulation **32** (1998) 53–56
19. Kokaram, A., Godsill, S.J.: Mcmc for joint noise reduction and missing data treatment in degraded video. IEEE Trans. on Signal Processing, Special Issue on MCMC **50** (2002) 189–205
20. Wexler, Y., Shechtman, E., Irani, M.: Space-time video completion. proc. of CVPR'04 **1** (2004) 120–127

Gaussian-Based Codebook Model for Video Background Subtraction

Yongbin Li, Feng Chen, Wenli Xu, and Youtian Du

Department of Automation, Tsinghua University, Beijing 100084, China
liyb04@mails.tsinghua.edu.cn,
{chenfeng, xuwl}@mail.tsinghua.edu.cn,
dyt02@mails.tsinghua.edu.cn

Abstract. As an effective method of background subtraction, codebook model suffers from unacceptable false negative detection rate in many situations due to its quantization criterion. In this paper, we propose an improved codebook model to solve this problem. Instead of using the original quantization criterion, we quantize the temporal series of the observations at a given pixel into codewords based on the Gaussian distribution assumption. We have performed this approach in our surveillance system for outdoor scenes and achieved excellent detection results.

1 Introduction

The segmentation of moving objects in videos is a fundamental task in many computer vision applications, such as intelligent video surveillance. A common approach to this is background subtraction. Its idea is to subtract the current frame from a reference background model.

If the scene background is static, the pixel values over time can be reasonably modeled with a Gaussian distribution. However, the natural scenes are not static in most situations. There are many non-static objects like waving trees, flickering monitors, escalators in the background. To solve these bimodal/multimodal backgrounds, Stauffer et al. [1] proposed the pixel-wise Gaussian mixture model (GMM). Power and Schoonees [2] expounded Stauffer's algorithm in terms of the theoretical foundation. Many authors have proposed improvements to this algorithm, including [3], [4]. Although GMM has many advantages, it does have disadvantages. Background with high frequency variations is not easily modeled by a few Gaussian components accurately. Moreover, using constant number of Gaussian components for every pixel is not necessary because many pixels are unimodal and only need one Gaussian distribution to fit them.

Recently, Kim et al. [5], [6] proposed a quantization/clustering technique called codebook model to construct a non-parametric background model, which bears the most similarity to our work. According to their clustering criterion, some pixels, which belong to the foreground objects but have the similar color to the shadows or highlights, are always misclassified as the background. The false negative rate can not be accepted in many situations.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 762–765, 2006.

We describe the improved algorithm procedure in great detail in section 2. In section 3, the experiment results and analyses will be shown. Finally, conclusions are presented in section 4.

2 Gaussian-Based Codebook Model

2.1 Background Modeling and Foreground Segmentation

This approach is a pixel-wise modeling technique so we ignore the pixel index for convenience. Let $X = \{x_1, x_2, ..., x_N\}$ be the training sequence for a given pixel consisting of N RGB vectors. Let $\ell = \{c_1, ..., c_L\}$ represent the codebook for this pixel consisting of L codewords. Each codeword c_i, $i = 1, 2, ..., L$, comprises a mean vector $\mu_i = (\bar{R}_i, \bar{G}_i, \bar{R}_i)$, a diagonal covariance matrix Σ_i where the diagonal elements are $\sigma^2_{i,R}$, $\sigma^2_{i,G}$, $\sigma^2_{i,B}$, and a 3-tuple $aug_i = (f_i, \lambda_i, q_i)$ where f_i records the frequency that the codeword has occurred, λ_i is defined as the maximum interval of time that the codeword has not recurred, and q_i is the last access time. Let M denote the final background model and the algorithm procedure is shown in Table 1.

For a new coming frame, each pixel is compared with the codewords of the corresponding codebook in the background model M. If there is one codeword matched, then the pixel is labeled as background and the codeword is updated. Otherwise, the pixel is labeled as foreground.

2.2 Background Model Update

The background model above can adapt to the slow illumination changes. But there exist other changes including parked cars and new appearing period-like

Table 1. Algorithm for Background Model Construction

I. $L \leftarrow 0$, $\ell \leftarrow \emptyset$;
II. For $x_t = (R_t, G_t, B_t)^T$ $(t = 1\ to\ N)$ do
  1) Find the codeword c_m in $\ell = \{c_i | 1 \le i \le L\}$ matching to x_t based on the following condition: $x_t^T \sum_m^{-1} x_t < T^2$ (usually $T = 3$)
  2) If $\ell = \emptyset$ or there is no match, then $L \leftarrow L + 1$. Create a new codeword c_L by setting:
    a) $\mu_L \leftarrow (R_t, G_t, B_t)$;
    b) $\sigma^2_{L,Z} = const \quad Z \in \{R, G, B\}$ (const usually ranges from 100 to 400);
    c) $aug_L \leftarrow (1, t-1, t)$;
  3) Otherwise, update the matched codeword c_m by setting:
    a) $\mu_m \leftarrow \alpha x_t + (1-\alpha)\mu_m$; ($\alpha \in$ (0,1), usually ranges from 0.005 to 0.01)
    b) $\sigma^2_{m,Z} \leftarrow (1-\alpha)\sigma^2_{m,Z} + \alpha(x_{t,Z} - \mu_{m,Z})(x_{t,Z} - \mu_{m,Z}) \quad Z \in \{R, G, B\}$;
    c) $aug_m \leftarrow (f_m + 1, \max\{\lambda_m, t - q_m\}, t)$;
  End for;
III. For each codeword c_i, $i = 1, ..., L$, calculate $\lambda_i \leftarrow \max\{\lambda_i, N - q_i\}$;
IV. $M = \{c_m | c_m \in \ell, \lambda_m < T_M\}$ (Usually T_M is equal to $N/2$).

Table 2. Algorithm for Background Model Update

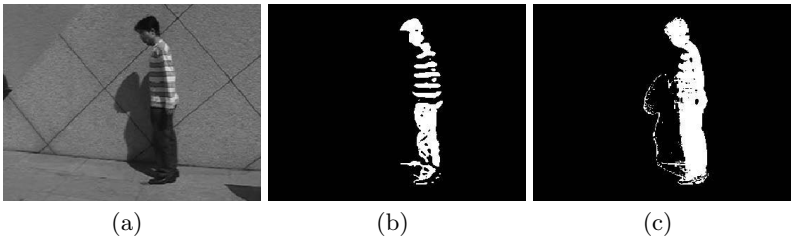
I. After training, M is obtained. Create a new model H as a cache.
II. For a new coming pixel, find a matched codeword in M. If found, update the codeword.
III. Otherwise, try to find a matched codeword in H and update it. If no match, create a new codeword h and add it to H.
IV. Filter out the cache codewords based on T_H. $H \leftarrow H - \{h_i \| h_i \in H, \lambda_i > T_H\}$
V. Move the cache codewords staying for enough time to M. $M \leftarrow M \cup \{h_i \| h_i \in H, f_i > T_{add}\}$
VI. Delete the codewords not accessed for a long time from M. $M \leftarrow M - \{c_i \| c_i \in M, \lambda_i > T_{del}\}$
VII. Repeat the process from step II

motions. In order to adapt to these changes, we build a cache model to update the background. The construction of the cache model is similar to the background model. The update procedure is shown in Table 2.

3 Experiments and Analyses

In the original method [5], [6], the shadows and the highlights are classified as background objects simultaneously. This is not feasible in many natural scenes. As shown in Fig. 1(b), the person is detected into pieces with the original method because some pixels which belong to the person have the similar property to that of the shadows and are easily misclassified as background objects. In our approach, we first detect the shadow as foreground objects and then implement the color constancy to verify them. Fig 1(c) shows the result of our approach.

In Fig. 2(a), there is a swaying tree in the scene. Fig. 2(b) shows the GMM detection result and Fig. 2(c) shows ours. They both achieved good performance. However, GMM needs to set 3 components for every pixel, while our method only needs 1.81 codewords per pixel. In most situations, our approach can get a more compact representation of the background compared with GMM.



(a) (b) (c)

Fig. 1. The detection results. (a) the original frame; (b) the detection result of the original codebook model [5], [6]; (c) the detection result of our approach.

(a) (b) (c)

Fig. 2. Comparison between GMM and our approach. (a) the original frame; (b) the detection result with GMM; (c) the detection result with our approach.

In our approach, the detection results are decided by the parameters T_M, T_H, T_{add} and T_{del}. The larger T_M is, the more codewords will be in the background model. T_H is similar to T_M and it is used for the cache model. The function of T_{add} and T_{del} is like that of the learning rate in GMM. All of them need to be determined according to the scene. In the above two experiments, $T_M = N/2$, $T_H = 50$, $T_{add} = 400$ and $T_{del} = 200$.

4 Conclusions

In this paper, we solve the problem that the false negative detection rate is too large to be accepted in many situations in Kim's codebook model. A comparison has also been made between our improved codebook model with GMM. The results show that our approach can achieve a more compact representation of the background without performance reduction.

References

1. Stauffer, C., Grimson, W.: Adaptive Background Mixture Models for Real-Time Tracking. In Proc. CVPR (1999) 246-252
2. Power, P.W., Schoonees, J.A.: Understanding Background Mixture Models for Foreground Segmentation. Proc. Image and Vision Computing New Zealand (2002)
3. KaewTraKulPong, P., Bowden, R.: An Improved Adaptive Background Mixture Model for Real-Time Tracking with Shadow Detection. Proc. European Workshop Advanced Video Based Surveillance Systems. (2001)
4. Lee, D.S.: Effective Gaussian Mixture Learning for Video Background Subtraction. IEEE Trans. PAMI 27 (2005) 827-832
5. Kim, K., Chalidabhongse, T.H., Harwood, D., Davis, L.: Background modeling and subtraction by codebook construction. International Conference on Image Processing (2004)
6. Kim, K., Chalidabhongse, T.H., Harwood, D., Davis, L.: Real-time Foreground - Background Segmentation using Codebook Model. Real-Time Imaging 11 (2005) 172-185

Frequency Domain Volume Rendering Based on Wavelet Transformation*

Ailing Ding[1,2] and Qinwu Zhou[3]

[1]Institute of Intelligence Information Processing,
Xidian University, Xi'an, 710071, China
alding@chd.edu.cn
[2] School of Information Engineering, Chang'an University,
Xi'an 710064, China
[3] School of Life Science and Technology, Xi'an Jiaotong University,
Xi'an 710049, China

Abstract. A new Fourier-wavelet volume rendering algorithm (FWVR) in frequency domain is presented and corresponding preprocessing is given in this paper. After the wavelet transforming and obtaining slice plane, FWVR finish the volume rendering image with a wavelet-based X-ray rending method in the frequency domain. FWVR comprises and unifies the advantages of recently introduced Fourier domain volume rendering techniques and wavelet based volume rendering. For the algorithm is based on the CT imaging and wavelet transformation, the time complexity of this method is $O(N^2 \log N)$ which is reduced apparently. Experiment of medicine volume data shows that this method is efficient. Finally, the potential application of FWVR is introduced in network rendering for saving transmission time.

1 Introduction

The medical volumetric data with some resolutions is data scene based regular grid. 3-D visualization has three methods: surface rendering, volume rendering and hybrid rendering. Recently, because of the advantage of processing of volumetric data and performance in character information, volume rendering is widely used in medical area[1-3]. A Fourier-wavelet volume rendering (FWVR) is presented in this paper, which contributes to the solution of the above problems by significantly reducing the time complexity. FWVR obtain a 2-D image in view plane based on the principle of 3-D data, which improves the efficiency of remote network transmission.

2 Volume Rendering in Frequency Domain Based on Wavelet Transformation

Wavelet-based volume rendering[4] yields 3-D image by getting multi-resolution of volumetric data after the 3-D discrete wavelet transform. Presently, there are two

* Nation Natural Science Fund of China Under Grant No. 60372047.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 766 – 769, 2006.

methods for volume rendering based on wavelet: ray-tracing and ray-footprints. The former has a disadvantage of big complexity of computation. The later has the advantage of little time complexity but it can not hold absorbing characteristic of illumination.

2.1 Fast Wavelet Transformation in Frequency Domain

Since the volume rendering is completed in frequency domain, the wavelet transform needed to be computed in frequency domain. Let the Z-transform of filters hh、hg、gh、gg be HH , HG , GH , GG respectively, where $\mathrm{H}(z):=\sum_n h_n z^{-n}$. Four filters are as follows:

$$\begin{cases} \mathrm{HH}(z_1,z_2)=\mathrm{H}(z_1)\mathrm{H}(z_2) & \mathrm{HG}(z_1,z_2)=\mathrm{H}(z_1)\mathrm{G}(z_2) \\ \mathrm{GH}(z_1,z_2)=\mathrm{G}(z_1)\mathrm{H}(z_2) & \mathrm{GG}(z_1,z_2)=\mathrm{G}(z_1)\mathrm{G}(z_2) \end{cases} \tag{1}$$

The decomposition in frequency domain is given by :

$$\begin{aligned} \mathrm{C}^{j+1}(z_1,z_2)&=\left[\tilde{\mathrm{H}}\tilde{\mathrm{H}}\cdot\mathrm{C}^{j}\right]^{down}(z_1,z_2) \quad \mathrm{D}^{j+1,1}(z_1,z_2)=\left[\tilde{\mathrm{H}}\tilde{\mathrm{G}}\cdot\mathrm{C}^{j}\right]^{down}(z_1,z_2) \\ \mathrm{D}^{j+1,2}(z_1,z_2)&=\left[\tilde{\mathrm{G}}\tilde{\mathrm{H}}\cdot\mathrm{C}^{j}\right]^{down}(z_1,z_2) \quad \mathrm{D}^{j+1,3}(z_1,z_2)=\left[\tilde{\mathrm{G}}\tilde{\mathrm{G}}\cdot\mathrm{C}^{j}\right]^{down}(z_1,z_2) \end{aligned} \tag{2}$$

Formula (1) and (2) are Fourier-wavelet decomposition (FWD) and Fourier-wavelet reconstruction (FWR) respectively.

2.2 Wavelet X-ray Transformation

The wavelet X-ray transformation[4] for volume rendering is proposed, followed by a procedure derived for wavelet transform in frequency domain. After the preprocessing by ordinary FVR, the wavelet decomposition of slice plane in frequency domain is performed.

The discrete *wavelet X-ray transform* in a 2-D wavelet series is defined as follows:

$$\mathrm{P}_\theta\, f(u,v)=\sum_{k,l} c_{k,l}^{M}(\theta)\Phi_{M,k,l}^{0}(u,v)+\sum_{j=1}^{M}\sum_{\tau=1}^{3}\sum_{k,l} d_{k,l}^{j,\tau}(\theta)\Psi_{j,k,l}^{\tau}(u,v) \tag{3}$$

,

where wavelet coefficients $c_{k,l}^{M}$, $d_{k,l}^{j,\tau}$ depend on the view angle θ, M denotes the depth of the pyramid. The coefficients of the formula (11) can be computed as follows:

Theorem 1. Let $f\in L^2(R^3)$, then coefficients in formula (3) for X-ray transform are:

$$c_{k,l}^{M}(\theta)=F_2^{-1}\left(F_2 P_\theta f\cdot F_2\tilde{\Phi}_M'^{0}\right)\left(2^j k,2^j l\right), \tag{4}$$

$$d_{k,l}^{j,\tau}(\theta)=F_2^{-1}\left(F_2 P_\theta f\cdot F_2\tilde{\Psi}_j'^{\tau}\right)\left(2^j k,2^j l\right). \tag{5}$$

where $\tilde{\Phi}'^{0}_{M}(u,v) = \tilde{\Phi}^{0,*}_{M,0,0}(-u,-v)$ and $\tilde{\Psi}'^{\tau}_{j}(u,v) = \tilde{\Psi}^{\tau,*}_{j,0,0}(-u,-v)$. $F_2(\ )$ and $F_2^{-1}(\ )$ denote the Fourier transform and inverse Fourier transform respectively. ~ and * denote the dual and complex conjugate respectively.

A client-sever visualization system is implemented by this method. The sever finishes the initial 3-D Fourier transform. Both the approximation and detail coefficients are sent to client. After clients perform the Fourier-wavelet reconstruction and IFFT, an approximation image is obtained.

2.3 Time Complexity

Assume that the input image C^0 is square and contains N^2 elements, where N is a power of two. Then the maximal number of decomposition levels is $M = \log_2 N$. Since each coefficient's computation requires at 4 additions, $4(N/2^j)^2$ multiplications for the jst decomposition level are needed. As a result, the computational complexity in the frequency domain is $O(N^2)$. Since the complexity of the FFT is $O(N^2 \log_2 N)$ and the reverse transform can be computed in the same time, the overall time complexity of FWVR is obviously very favorable: $O(N^2 \log_2 N)$.

3 Experiments

Experiments with CT data of size 128^3 were carried out for Bior5.5 wavelets of Biorthogonal wavelet basis. The volume rendering is implemented by FWVR. Fig.1 shows the reconstruction from Bior5.5 wavelet decomposition (decomposition level is M=3/2/10) for volume rendering of gray image. FWVR costs the 6.18 seconds, while wavelet splatting has timings of 12.62s with the same sets and image resolution . Fig.2 shows high-quality pseudo-color image rendered by FWVR with the same wavelets and decomposition level. Experiments shows that scaling of the timings for increasing data size is in agreement with the theoretical complexity.

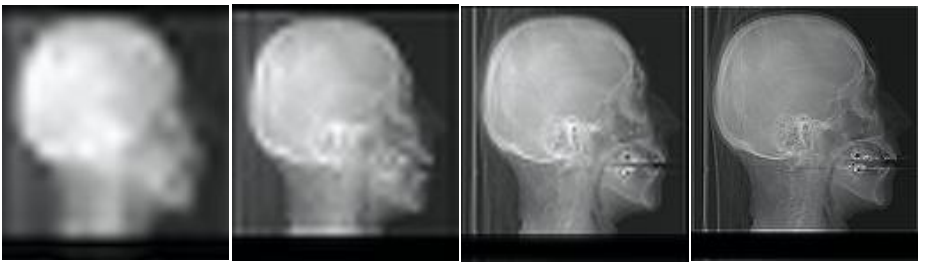


Fig. 1. volume rendering of gray image in frequency domain

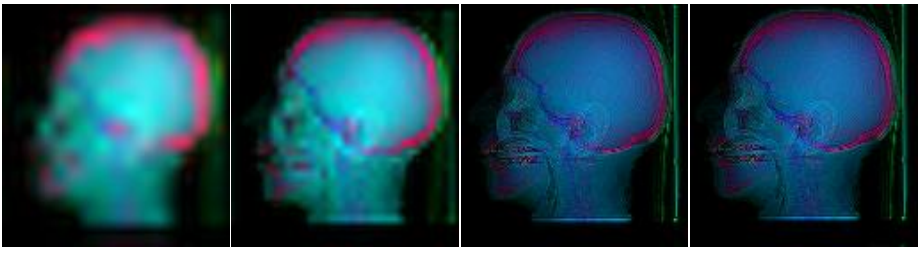


Fig. 2. Volume rendering of pseudo-color image in frequency domain

4 Conclusions

The volume rendering in frequency domain based on the wavelet transform is presented in this paper, and a new Fourier-wavelet volume rendering algorithm (FWVR) in frequency domain is given here. This method has the advantage of the small time complexity, which is $O(N^2 \log N)$, over the other algorithm in spatial domain. The complexity of the computation is reduced with no loss in the performance and to provide comparable results. In the coming versions of the wavelet based rendering system, web-based user interfaces and servers that support simultaneous access from various sites will be implemented. The system will be integrated with a network module (client/server) in future versions.

References

1. Keppel E.: Approximating complex surfaces by triangulation of contour lines. IBM Journal of Research Development, 1 (1975)2-11.
2. Lorensen W., Cline H. E.: Marching Cubes: A high resolution 3D surface construction algorithm. Computer Graphics, , 4 (1987) 163-169.
3. Levoy M.: Volume Rendering: Display of surfaces from volume data. IEEE Computer Graphics & Application, 5(1998): 29-37.
4. R. A. Zuidwijk, “The wavelet X-ray transform,” Centre Math. Com-puter. Sci., Amsterdam, The Netherlands, Tech. Rep. PNA-R9703, Mar. 1997.

A New Method for Compression of SAR Imagery Based on MARMA Model*

Jian Ji[1], Zheng Tian[2], and Yanwei Ju[2]

[1] Department of Computer Science & Technology, Northwestern Polytechnical University, Xi'an, 710072, China
jijiangao@gmail.com
[2] Department of Applied Mathematics, Northwestern Polytechnical University, Xi'an, 710072, China

Abstract. In this paper, we present a new method of SAR imagery compression based on multiscale autoregressive moving average (MARMA) models. We use the multiscale representation as the cornerstone of the modeling process, and construct the MARMA models of SAR image. We then predict the initialized image data using these multiscale models. Next we compress image data through coding the residual image. Extension simulations have proven that the proposed method achieves high compression radios with impressive image quality.

1 Introduction

Synthetic aperture radar (SAR) data represents an important source of information for a large variety of scientists around the world. However, while the volume of data collected is increasing as fast, the ability of transmit it to the ground, or to store it, is not increasing as fast. Thus, there is a strong interest in developing data encoding and decoding algorithms which can obtain higher compression ratios while keeping image quality to an acceptable level [1]. Recently, different multiscale algorithms have been proposed to SAR image analysis [2-6]. The MAR model has been applied to the problem of compression in SAR image by Andrew. Kim [5]. That method compresses SAR image based on the results of supervised segmentation, and needs the prior knowledge for SAR image segmentation. In [6], MARMA processes were defined and applied to classification for SAR image. But it didn't take further research on theory and other applications for MARMA models. In this paper, we present an approach to construct MARMA modeling frameworks, and applied them to SAR image compression. Our compression method doesn't require the prior knowledge for supervised segmentation. We predict the SAR image data for compression using these multiscale models and compress image data through coding the residual image.

* This work is supported by the National Natural Science Foundation of China (60375003), Graduate starting seed fund of Northwestern Polytechnical University (Z200566).

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 770 – 773, 2006.

2 MARMA Model for SAR Imagery

Assuming I is the SAR image for compression, the starting point for our model development is a multiscale sequence $I_0 \cdots I_R$ of SAR images I, where I_0 and I_L correspond to the finest and coarsest resolution images, respectively. Specifically, we assume that I_0 is equal to I. Each pixel $I_m(k,l)$ in I_m is obtained by

$$I_m^{'}(k,l) = 20\log_{10}\left|\sum_{i=2^m k}^{2^m(k+1)-1}\sum_{j=2^m l}^{2^m(l+1)-1} I(i,j)\right|, \qquad I_m(k,l) = I_m^{'}(k,l) - C_m \tag{1}$$

where C_m is mean value of $I_m^{'}(k,l)$ and $m = 1 \ldots R$. This indicates that a quadtree is natural for the mapping, and each node s on the tree is associated with one of the pixels $I_m(k,l)$ corresponding to pixel (k,l) of SAR image I_m.

A MAR process defined on node s satisfies

$$I(s) = a_{1,m(s)} I(s\overline{\gamma}) + \ldots + a_{R,m(s)} I(s\overline{\gamma}^R) + \omega(s) \tag{2}$$

where R is the order of autoregression, and $a_{m(s)} = (a_{1,m(s)}, \ldots, a_{R,m(s)})^T$ are the real, scalar-valued regression coefficients. The term $\omega(s)$ is a white noise process which can be viewed as the modeling error image.

Furthermore, we develop a multiscale sequence $w_0, w_1, \ldots, w_Q$ of the MAR model error image $\omega(s)$. Then we define a MARMA process on a node s like so:

$$I(s) = a_{1,m(s)} I(s\overline{\gamma}) + \ldots + a_{R,m(s)} I(s\overline{\gamma}^R) + b_{1,m(s)} w(s\overline{\gamma}) + \ldots + b_{Q,m(s)} w(s\overline{\gamma}^Q) + \xi(s) \tag{3}$$

here $b_{1,m(s)}, \ldots, b_{Q,m(s)}$ are coefficients of a multiscale moving average process, R and Q are orders of autoregression and moving average respectively. The term $\zeta(s)$ is white-noise and can be viewed as residual image.

3 Compression of SAR Imagery Based on MARMA Model

Assuming that the coarser resolutions $I_1 \cdots I_R$, $w_1 \cdots w_Q$ have already been coded to $\tilde{I}_1 \cdots \tilde{I}_R$, $\tilde{w}_1 \cdots \tilde{w}_Q$, we compute each node s in the predicted image $\hat{I}_0$ from Eq. (3) neglecting the noise term $\zeta(s)$. The residual image is then computed by Eq. (4)

$$E_0 = I_0 - \hat{I}_0 \tag{4}$$

Because E_0 has a much smaller dynamic range than I_0 and conveys all the necessary information for decoder to reconstruct I_0, E_0 can be coded instead of I_0. To further improve compression performance, residual image E_0 is transformed using discrete wavelet transform (DWT) and a soft threshold is applied to all the transformed coefficients. Following the soft threshold procedure, the coefficients are uniformly quantized and entropy coded. Then data of SAR image compression is received. The SAR image compression and decompression methods are shown in Fig. 1. The structure within the long dashed boundary and the short dashed boundary correspond to compression procedures and decompression procedures respectively.

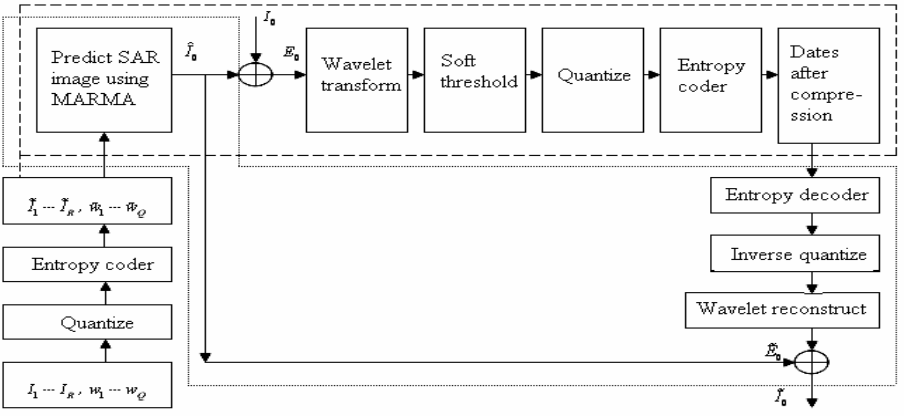


Fig. 1. Compression and decompression methods of SAR imagery based on MARMA models

Because MAR models are a special case of MARMA models, the compression method based on MAR models is a special case of the MARMA compression method in our paper. The difference between MAR and MARMA method is that we use MAR models when we predict original image I_0 in MAR compression method.

4 Experimental Results

In order to evaluate the performance of our approaches to SAR image compression, we have applied them to a SAR image with size of 256x256 pixels as Fig. 2(a).

Three different models were fitted to the SAR data:

- a second order autoregression dubbed MAR(2):

$$I(s) = -0.050888 I(s\overline{\gamma}) + 0.44427 I(s\overline{\gamma}^2) + \omega(s) \tag{5}$$

- a fourth order autoregression MAR(4):

$$I(s) = -0.021294 I(s\overline{\gamma}) + 0.1399 I(s\overline{\gamma}^2) + 0.41297 I(s\overline{\gamma}^3) + 0.4321 I(s\overline{\gamma}^4) + \omega(s) \tag{6}$$

- a second order autoregressive moving average dubbed MARMA(2,2):

$$I(s) = -.05089 I(s\overline{\gamma}) + 0.4443 I(s\overline{\gamma}^2) - 0.01379 w(s\overline{\gamma}) + 0.04138 w(s\overline{\gamma}^2) + \xi(s) \tag{7}$$

Fig.2 shows how the compression performance of our method compares with that of the JPEG algorithms. In order to evaluate the quality of compression images, we use the peak signal-to-noise ratios (PSNR's) of the reconstructed images in (8)

$$\mathrm{PSNR} = 10\log_{10}\left[MN(\max_{[m,n]}\{\hat{f}(m,n)\} - \min_{[m,n]}\{\hat{f}(m,n)\})^2 \Big/ \sum_{m=1}^{M}\sum_{n=1}^{N}\left[\hat{f}(m,n) - f(m,n)\right]^2 \right] \tag{8}$$

Where M and N are the dimensions of the finest resolution image, $f(m,n)$ is the finest resolution image, and $\hat{f}(m,n)$ is the reconstructed image. Table 1 displays the compress radios and PSNR of the reconstructed images. The extent to which JPEG could

compress the images was limited, hence, the reason for they're being larger than the MAR (2), MAR (4) and MARMA (2, 2) coded images. Even so, JPEG still performed considerably worse than the other three. The MAR (2), MAR (4) and MARMA (2, 2) methods perform very well at this high compression radio. Among the MAR (2), MAR (4) and MARMA (2, 2), MAR (2) has highest compression radio and lowest PSNR, MARMA (2, 2) has lowest compression radio and highest PSNR.

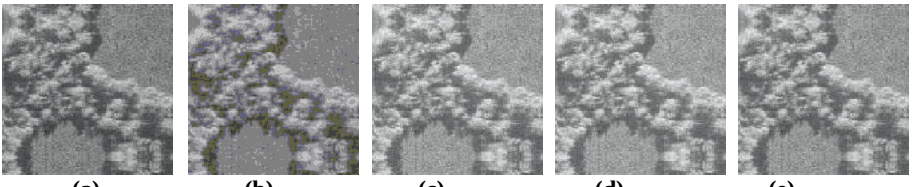


Fig. 2. Compression results: **(a)** initial SAR image composed of grass and forest; **(b)** JPEG 5079 byte; **(c)** MAR (2) 1580 byte; **(d)** MAR (4) 1770 byte; **(e)** MARMA (2, 2) 1919 byte

Table 1. PSNR comparing JPEG, MAR (2), MAR (4) and MARMA (2, 2)

Algorithms	JPEG	MAR (2)	MAR (4)	MAR(2, 2)
Bits/pixel	0.0775	0.0241	0.027	0.0293
PSNR	28.5dB	29.66dB	29.12dB	31.99dB

5 Conclusion

We have presented a new method of SAR imagery compression based on MARMA models. The result proved that MARMA models could be developed to compress SAR image, and achieve high compression radios with impressive image quality.

References

1. Benz, U., Strodl, K. and Moreria, A.: A comparison of several algorithms for SAR raw data compression. IEEE Trans. Geos. Remote Sense. Vol.33, no.5, (1995) 1266-1276.
2. 2 Xian-Bin Wen, Zheng Tian.:Mixture multiscale autoregressive modelling of SAR imagery for segmentation. Electronics Letters. 39(17), (2003)1272–1274.
3. Fosgate, C.H., Krim, H., Irving, W., Karl, W. C. and Willsky, A. S.: multiscale segmentation and anomaly enhancement of SAR imagery. IEEE Trans. Image Processing. vol.6, no.1. (1997) 7-20.
4. H Choi, R G Baraniuk.: Multiscale image segmentation using wavelet-domain hidden Markov models. IEEE Trans. Image Processing, 2001, 10: 1309-1321.
5. Andrew. Kim and Hamid.Krim.: hierarchical stochastic modeling of SAR imagery for segmentation/compression. IEEE Trans Processing. vol. 47, no.2. , (1999) 458-468.
6. Schoeder, J., Howard, D. and Guanawardean, A.: multiscale modeling for target detection in complex synthetic aperture radar imagery. In Proceedings of the Asilomar Conference on Signals, Speech and Signal Processing, Phoenix, AZ. (1999) 77-82.

Geometrical Fitting of Missing Data for Shape from Motion Under Noise Distribution

Sungshik Koh[1] and Chung Hwa Kim[2]

[1] Insan Innovation Telecom Co., Ltd.,
524-19, Unnam-dong, Gwangsan-gu, Gwnagju, 506-812, Korea
phdkss@chosun.ac.kr
[2] Dept. of Electronics, Chosun University,
375, Seosuk-dong, Dong-gu, Gwangju, 501-759, Korea
jhdkim@chosun.ac.kr

Abstract. When converting image sequence to 3D, several entries of the matrix have not been observed by occlusions and other entries have been perturbed by the influence of noise. In this paper, we propose a method for fitting geometrically missing data in noisy observation matrix with iterative SVD factorization. The main idea of the proposed algorithm is that the orientation and distance of noisy vector ca be handled directly by geometrical properties between 2D image plane and 3D error space under noise distribution. To confirm the recoverability of missing data, we carry out the experiments for synthetic and real sequences. The results in practical situations demonstrated with synthetic and real video sequences verify the efficiency and flexibility of the proposed method.

1 Introduction

In computer vision, matrix factorization methods have been widely used to reconstruct 3D shape and camera motion from 2D image data matrices. To be factored out, it is essential that all of Feature Points (FPs) are visible in the measurement matrix. In order to provide sufficient 3D content, it is important to convert existing 2D video footage into 3D data [1], [2]. However, in practice, the observation matrix may be incomplete, i.e., some of entries may be unobserved. Therefore, there are actually some weak points in matrix factorization, resulting from tracking failure.

Many researchers [3]-[12] have developed 3D reconstruction methods using SVD factorization methods in difference ways. The matrix collects 2D trajectories of projections of FPs [3]-[8] or other primitives [9]-[12]. In real life video clips, these projections are not visible along the entire image sequence due to occlusion and a limited field of view. Thus, the observation matrix is incomplete. Sub-optimal solutions were proposed in [3] and [13]. In [3], the missing values of the observation matrix are 'filled in', in a sequential and heuristic way, using SVD of observed partial matrices. Filling of missing data was first realized by Tomasi and Kanade [3] for affine camera. Jacobs [13] improved their method by fitting an unknown matrix of a certain rank to an incomplete noisy matrix resulting from measurements in images,

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 774 – 783, 2006.

which is called Linear Fitting. However, his method presents no approach to determine whether the incomplete matrix is stable or unstable. Guerreiro and Aguiar [14] proposed Expectation-Maximization (EM) and Two-Step (TS) iterative algorithm. The algorithms converged to the global optimum in a very small number of iterations. However, the performance of both EM and TS are influenced by noisy observation.

In the paper, a new estimation method that is executed under a noisy observation matrix is proposed. In particular, geometrical correlations on the proposed algorithm's behavior are studied, and compared with EM and TS. The experimental results demonstrate that the proposed algorithm is more robust than EM and TS, with respect to the sensibility to noise. An affine camera model is mainly treated in this paper, but the method can be extended to a perspective camera model because of handling missing data using geometrical correlations. The remainder of this paper is composed as follows. In Section 2, the relationship between the number of frames and 3D shape error is introduced. Section 3 describes the estimation algorithm of missing data. In Section 4, the behavior of the proposed algorithm is illustrated with synthetic and real video sequences that demonstrate good recoverability when dealing with arbitrary matrices. Finally, Section 5 concludes this paper.

2 3D Shape Error by the Number of Image Frames

When the observation matrix has no missing data, the error of the reconstructed 3D shape is not related to the number of image frames. However, in practice, several entries of the matrix may not be observed, or observed with noise, due to image resolution, occlusion and so on. As a result, the 3D shape error is relative to the number of available image frames. Although most researchers [15]-[19] have analyzed noise characteristics in detail, the problem of estimating deficient matrices from noisy observations of their entries has no closed form solution. In addition, it is difficult to evaluate the reconstructed 3D shape under noise distribution accurately, even though it is perturbed by just one noisy entry in the observation matrix. The relationship between the number of frames and 3D shape error are now investigated. Given an observation matrix $\boldsymbol{W}$ of a *M*x*N* incomplete matrix $\tilde{\boldsymbol{W}}$ corrupted by white Gaussian noise, its *Root Mean Square* (*RMS*) error is

$$RMS = \sqrt{\frac{\sum_{i=1}^{M\times N}\left\|\boldsymbol{W}-\tilde{\boldsymbol{W}}\right\|_F^2}{M\times N}}, \tag{1}$$

where $\|\cdot\|_F$ means the Frobenius norm. Given a true point ($\boldsymbol{P}^i$) and a perturbed point ($\tilde{\boldsymbol{P}}^i$) reconstructed from the corrupted FP, the *position Root Mean Square* (*pRMS*) error is defined in order to evaluate the 3D shape error.

$$pRMS = \underset{\boldsymbol{R},\boldsymbol{T}}{argmim}\sqrt{\frac{\sum_{n=1}^{N}\left\|\left(\boldsymbol{R}\cdot\tilde{\boldsymbol{P}}^n+\boldsymbol{T}\right)-\boldsymbol{P}^n\right\|_F^2}{N}}, \tag{2}$$

where N is the number of 3D points. $\boldsymbol{R}$ and $\boldsymbol{T}$ denote the rotation matrix and the translation vector, respectively.

3D reconstruction error is dependant on the total viewing angle and the total number of frames [18], [19]. That is, the *RMS* error of (1) is constant with respect to the number of image frames. However, the *pRMS* error of (2) is strongly influenced according to the number of image frames when the entries of the matrix are observed with noise. It means that, when the observation matrix has noisy entries, a greater number of image frames provide a better result. Therefore, in order to obtain accurate 3D contents it is obviously essential that the missing data of the observation matrix must be estimated accurately.

3 Geometrical Correlations Estimation Algorithm of Missing Data

In this section, some specific geometry between the noise of the 2D observation matrix and the error of 3D shape is described.

3.1 3D Error Space

It is extremely challenging to accurately evaluate the precision of the 3D shape reconstructed from a noisy observation matrix. In order to estimate and evaluate missing data, at first, 3D error space is introduced as the following:

i) For recovering a missing FP ($\boldsymbol{p}_m$), its initial position ($\boldsymbol{p}_t + \Delta\boldsymbol{e}$) is first fit roughly and three FPs ($\boldsymbol{p}_{bA}$, $\boldsymbol{p}_{bB}$, and $\boldsymbol{p}_{bC}$) are randomly selected, which are called bias FPs being neighbors of the missing FP ($\boldsymbol{p}_m$), on the same 2-D image plane. Next, new FPs ($\boldsymbol{q}_i$) are added, which are called Reference Points (RPs), on a circular pattern ($r = \Delta\boldsymbol{c}_i$) centering on the missing FP ($\boldsymbol{p}_m$). The aspects are shown in Fig. 1(a), where Π_2 is a reference plane composed of $\boldsymbol{q}_i$ on the 2-D image plane and $\overrightarrow{\boldsymbol{p}_m\boldsymbol{q}_i}$ is a reference vector (RV) composed of $\boldsymbol{p}_m$ and $\boldsymbol{q}_i$ on the 2-D image plane.

$$\boldsymbol{p}_m = \boldsymbol{p}_t + \Delta\boldsymbol{e}\,, \tag{3}$$

$$\boldsymbol{q}_i = \boldsymbol{p}_m + \Delta\boldsymbol{c}_i\,,\ i = 1,2,\cdots,Z\,, \tag{4}$$

where $\boldsymbol{p}_t$, $\Delta\boldsymbol{e}$, $\boldsymbol{q}_i$, and $\Delta\boldsymbol{c}_i$ are a true FP, a noise vector, RPs, and circle radius on the 2D image plane, respectively.

ii) Using affine SVD factorization, the roughly fitted FP ($\boldsymbol{p}_m$), three bias FPs ($\boldsymbol{p}_{bA}$, $\boldsymbol{p}_{bB}$, and $\boldsymbol{p}_{bC}$), and circular RPs ($\boldsymbol{q}_i$) are reconstructed to $\boldsymbol{P}_m^*$, ($\boldsymbol{P}_{bA}^*$, $\boldsymbol{P}_{bB}^*$, and $\boldsymbol{P}_{bC}^*$), and $\boldsymbol{Q}_i^*$ on the 3-D reconstruction space, respectively (see Fig. 1(b)).

$$\boldsymbol{Q}_i^* = \boldsymbol{P}_m^* + \Delta\boldsymbol{C}_i^*\,, \tag{5}$$

where $\boldsymbol{P}_m^*$ and $\Delta\boldsymbol{C}_i^*$ are the reconstructed RPs and its circular parameter on 3-D reconstruction space. The symbol '*' means a perturbation.

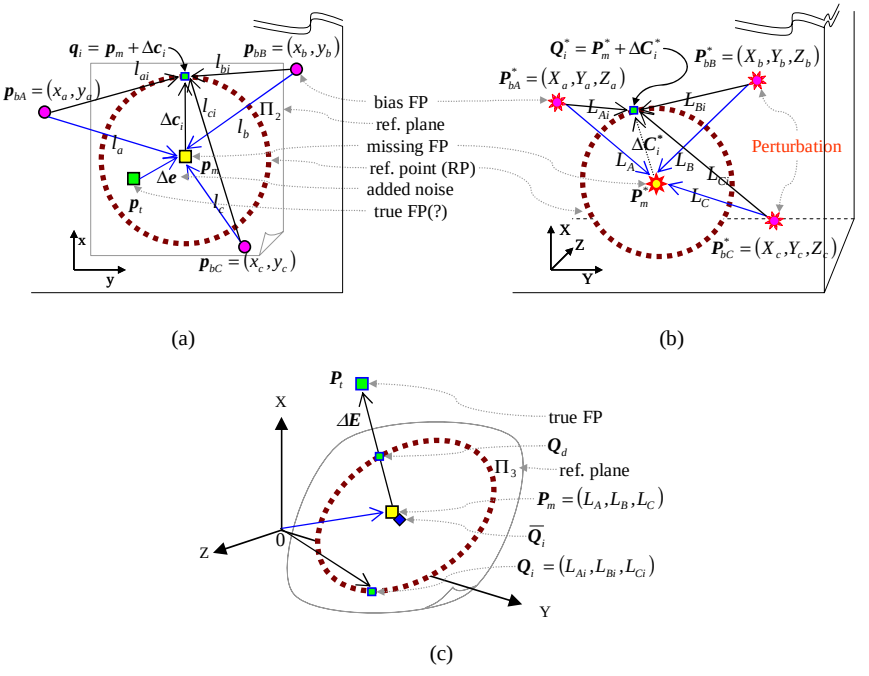


Fig. 1. Comparison of some parameters on 2D image plane, 3D reconstruction space, and 3D error space. (a) noisy FP, three bias FPs and circular RPs on 2D image plane. (b) parameters reconstructed from (a) on 3D reconstruction space. (c) parameters transformed from (b) on 3D error space.

iii) A 3-D error space is defined as the coordinates of 3-D point vectors without perturbation, which are transformed from three Euclidean distances between three bias FPs ($\boldsymbol{P}_{bA}^{*}$, $\boldsymbol{P}_{bB}^{*}$, and $\boldsymbol{P}_{bC}^{*}$) and a FP ($\boldsymbol{P}^{*}$) on 3-D reconstruction space. For example, the $\boldsymbol{P}_{m}$ on 3-D error space can be transformed from the missing FP ($\boldsymbol{P}_{m}^{*}$) on 3-D reconstruction space by three Euclidean distances between three bias FPs and the missing FP on the same 2-D image plane as

$$\boldsymbol{P}_{m} = (L_A, L_B, L_C), \tag{6}$$

where $L_A = \left\| \overrightarrow{\boldsymbol{P}_m^{*} \boldsymbol{P}_{bA}^{*}} \right\|$, $L_B = \left\| \overrightarrow{\boldsymbol{P}_m^{*} \boldsymbol{P}_{bB}^{*}} \right\|$, and $L_C = \left\| \overrightarrow{\boldsymbol{P}_m^{*} \boldsymbol{P}_{bC}^{*}} \right\|$. Also the circular RPs ($\boldsymbol{Q}_{\mathrm{i}}^{*}$) can be also expressed as $\boldsymbol{Q}_{i}$ on the 3-D error space as

$$\boldsymbol{Q}_{i} = (L_{Ai}, L_{Bi}, L_{Ci}), \tag{7}$$

where $L_{Ai} = \left\| \overrightarrow{\boldsymbol{P}_m^{*} \boldsymbol{P}_{bA}^{*}} \right\|$, $L_{Bi} = \left\| \overrightarrow{\boldsymbol{P}_m^{*} \boldsymbol{P}_{bB}^{*}} \right\|$, and $L_{Ci} = \left\| \overrightarrow{\boldsymbol{P}_m^{*} \boldsymbol{P}_{bC}^{*}} \right\|$, $i = \{1,2,\cdots,Z\}$.

This aspect is shown in Fig. 1(c), where Π_3 is an reference plane composed of $\boldsymbol{Q}_{i}$ on the 3-D error space and $\overrightarrow{\boldsymbol{P}_m \boldsymbol{Q}_i}$ is a Reference Vector (RV) composed of $\boldsymbol{P}_{m}$ and $\boldsymbol{Q}_{i}$ on the same 3-D error space.

3.2 Geometrical Correlations Between Π_2 and Π_3

In this section, the geometry between two reference planes Π_2 and Π_3 are analyzed on 2D image plane and 3D error space, respectively. In investigating the geometrical correlations, the following motivating facts are found (see Fig. 2):

***Plane*:** Since 2-D RVs are on a plane Π_2, the 3-D RVs are also approximately located on a plane Π_3.

$$\Pi_2 = \left\{\overrightarrow{\boldsymbol{p}_m\boldsymbol{q}_i},\ i=1,2,\ldots,Z\right\},$$
$$\Pi_3 \cong \left\{\overrightarrow{\boldsymbol{P}_m\boldsymbol{Q}_i},\ i=1,2,\ldots,Z\right\}.$$

***Pattern*:** If the RPs on Π_2 are distributed as a circular pattern, then the RPs on Π_3 are distributed as an ellipse and are very close to being circular.

$\{\boldsymbol{q}_i\}$: circular pattern on Π_2,

$\{\boldsymbol{Q}_i\}$: ellipse pattern on Π_3.

***Symmetry*:** If two of any RVs on Π_2 exist on symmetrical positions with $\boldsymbol{p}_m$, then their positions on Π_3 are nearly symmetric.

$$\overrightarrow{\boldsymbol{p}_m\boldsymbol{q}_1} \cong -\overrightarrow{\boldsymbol{p}_m\boldsymbol{q}_2} \Leftrightarrow \overrightarrow{\boldsymbol{P}_m\boldsymbol{Q}_1} \cong -\overrightarrow{\boldsymbol{P}_m\boldsymbol{Q}_2}.$$

***Size*:** If two of any RVs on Π_2 are in the same direction with different sizes, then the RVs on Π_3 keep their magnitude relationships and ratios relative to the size.

$$\left\|\overrightarrow{\boldsymbol{p}_m\boldsymbol{q}_1}\right\| < \left\|\overrightarrow{\boldsymbol{p}_m\boldsymbol{q}_2}\right\| \Leftrightarrow \left\|\overrightarrow{\boldsymbol{P}_m\boldsymbol{Q}_1}\right\| < \left\|\overrightarrow{\boldsymbol{P}_m\boldsymbol{Q}_2}\right\|, \quad \frac{\left\|\overrightarrow{\boldsymbol{p}_m\boldsymbol{q}_1}\right\|}{\left\|\overrightarrow{\boldsymbol{p}_m\boldsymbol{q}_2}\right\|} \cong \frac{\left\|\overrightarrow{\boldsymbol{P}_m\boldsymbol{Q}_1}\right\|}{\left\|\overrightarrow{\boldsymbol{P}_m\boldsymbol{Q}_2}\right\|}.$$

***Angle*:** If three RVs on Π_2 are arranged in some angles, then the RVs on Π_3 are also arranged similarly, while maintaining the relationship magnitude and ratio around the angle.

$$\angle\boldsymbol{q}_1\boldsymbol{p}_m\boldsymbol{q}_2 < \angle\boldsymbol{q}_1\boldsymbol{p}_m\boldsymbol{q}_3 \Leftrightarrow \angle\boldsymbol{Q}_1\boldsymbol{P}_m\boldsymbol{Q}_2 < \angle\boldsymbol{Q}_1\boldsymbol{P}_m\boldsymbol{Q}_3.$$

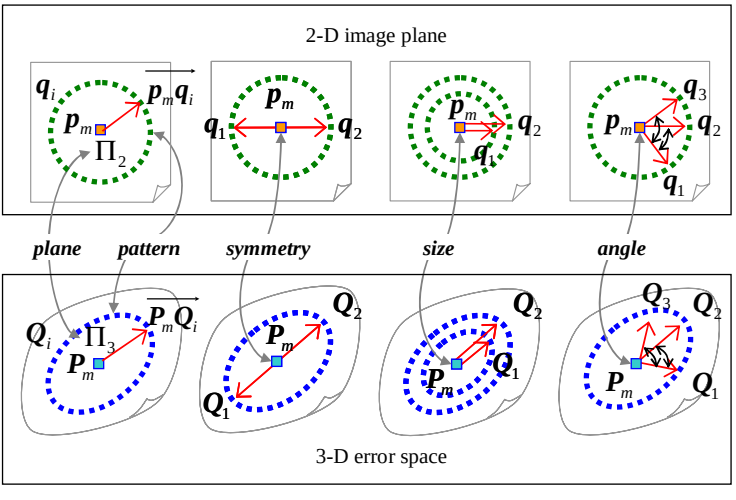


Fig. 2. Relationship of RVs between on Π_2 and Π_3

According to the above investigations, it can be observed that there are these geometrical correlations such as plane, pattern, symmetry, size, and angle between on Π_2 and Π_3. These aspects are always true for not only synthetic images but also real images. Therefore, the estimation algorithm for estimating the missing FPs can be derived using above facts in the next section.

4 Geometrical Fitting Algorithm

In this section, a new algorithm for solving the problem in estimating missing data from noisy observations of a subset of entries is introduced. The primary objective of this method is to obtain more accurate 3D reconstruction by estimating the missing data.

4.1 Geometrical Fitting of Missing Data

When a FP deviates from its observation matrix position, its 3-D point reconstructed by affine SVD factorization is also misaligned. In this section, we estimate the noise vector of the missing FP using the geometrical correlations described previously. Since $\boldsymbol{P}_m(L_A, L_B, L_C)$ and $\boldsymbol{Q}_i(L_{Ai}, L_{Bi}, L_{Ci})$ are transformed from (5) on the 3-D error space, an error vector of a missing FP can be expressed as

$$\Delta \boldsymbol{E} = \overrightarrow{\boldsymbol{P}_m \boldsymbol{P}_t}, \tag{8}$$

where $\boldsymbol{P}_t$ is the true FP on 3D error space. Because $\boldsymbol{P}_t$ is unknown parameter, an approximate $\boldsymbol{P}_t'$ is substituted. It can be obtained from a sub-matrix without missing FP. Also the relationship is satisfied with $\boldsymbol{P}_t' \cdot \overline{\boldsymbol{Q}}_i \cong \boldsymbol{P}_t \cdot \overline{\boldsymbol{Q}}_i'$, where $\overline{\boldsymbol{Q}}_i$ is the mean of $\boldsymbol{Q}_i$s. Hence, the approximate error vector can be represented as

$$\Delta \boldsymbol{E}' = \overrightarrow{\boldsymbol{P}_m \boldsymbol{P}_t'} = \boldsymbol{P}_t' \langle \times \rangle \left(\overline{\boldsymbol{Q}}_i \langle / \rangle \overline{\boldsymbol{Q}}_i' - (1,1,1) \right), \tag{9}$$

where $\boldsymbol{P} \langle \times \rangle \boldsymbol{Q} \equiv (x_1 x_2, y_1 y_2, z_1 z_2)$ and $\boldsymbol{P} \langle / \rangle \boldsymbol{Q} \equiv \left(\frac{x_1}{x_2}, \frac{y_1}{y_2}, \frac{z_1}{z_2} \right)$ for $\begin{cases} \boldsymbol{P} = (x_1, y_1, z_1) \\ \boldsymbol{Q} = (x_2, y_2, z_2) \end{cases}$.

If $\|\Delta \boldsymbol{E}'\| \neq 0$, this means some noise exists in $\boldsymbol{P}_m$ on 3D error space and also in $\boldsymbol{p}_m$ of (3). In order to obtain the noise vector ($\Delta \boldsymbol{e}'$) of the missing FP on Π_2, we first calculate the error vector from the parameters represented on Π_3 of the 3D error space.

$$\Theta_d = \underset{d \in i}{\boldsymbol{argmin}} \|\Psi_i\|^2, \quad for\ \Psi_i = \boldsymbol{cos}^{-1} \left(\frac{\left\| \overrightarrow{\boldsymbol{P}_m \boldsymbol{P}_t'} \right\| \cdot \left\| \overrightarrow{\boldsymbol{P}_m \boldsymbol{Q}_i} \right\|}{\overrightarrow{\boldsymbol{P}_m \boldsymbol{P}_t'} \cdot \overrightarrow{\boldsymbol{P}_m \boldsymbol{Q}_i}} \right), \tag{10}$$

$$A_d = \frac{\left\| \overrightarrow{\boldsymbol{P}_m \boldsymbol{P}_t'} \right\|}{\left\| \overrightarrow{\boldsymbol{P}_m \boldsymbol{Q}_d} \right\|}, \ d \in \{1,2,\cdots,Z\}, \tag{11}$$

where Θ_d is the minimum angle between $\overrightarrow{P_m P_t'}$ and $\overrightarrow{P_m Q_d}$, and A_d is the ratio of the size of $\overrightarrow{P_m P_t'}$ based on $\overrightarrow{P_m Q_d}$. Next, according to the geometrical correlations, the noise vector of the missing FP on Π_2 is derived from (10) and (11) as

$$\theta_d = \cos^{-1}\left(\frac{\left\|\overrightarrow{p_m q_d}\right\| \cdot \left\|\overrightarrow{p_m q_1}\right\|}{\overrightarrow{p_m q_d} \cdot \overrightarrow{p_m q_1}} \right), \tag{12}$$

$$\alpha_d = A_d \left\|\overrightarrow{p_m q_d}\right\|. \tag{13}$$

Hence, the missing FP (p_m) can be updated as

$$\tilde{p}_t = p_m - \Delta e', \tag{14}$$

where $\Delta e' \equiv f(\alpha_d, \theta_d)$, which is a vector with magnitude α_d and angle θ_d. If $\|\Delta E'\|$ is larger than the predefined threshold, then $\tilde{p}_t$ is set up to p_m and the above procedure is repeated until the position converses sufficiently close to the true position.

4.2 Estimation Procedure for Multi-frame

In order to accurately estimate the missing data from noisy observations of a subset of its entries, the missing data is required to be estimated until the noise level is included in the noisy observation matrix. In this paper, a minimum of three image frames with no missing data are required, which creates a sub-matrix. However, the other frames are allowed irrespective of missing data. This approach iteratively uses general affine SVD factorization to fit each missing data. It is explained in detail in Fig. 3. The estimation procedure begins from four frames which consist of a sub-matrix frame, and a frame with missing data. The size of the sub-matrix increases gradually after solving the missing data in each frame.

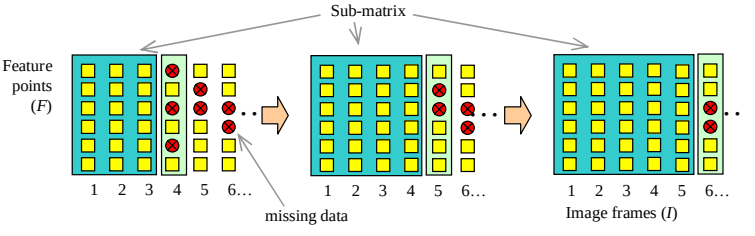
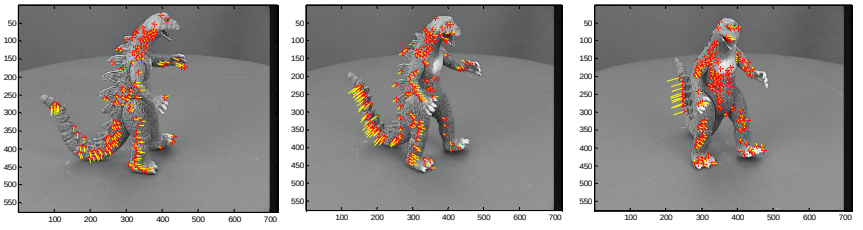


Fig. 3. Procedure for estimating missing data

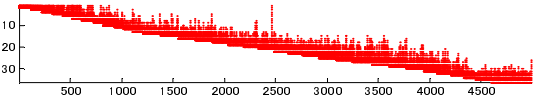
5 Experimental Results

In this section, we describe experiments that illustrate the behavior of the proposed algorithm using real video sequences. The proposed system is confirmed through experimental results by comparing with other approaches, EM and TS.

In order to test the proposed algorithm on real data, the entries of an observation matrix are tracked over the 36-frame 'dinosaur' sequence [20]. The 3rd, 6th, and 9th frame (720x576) are presented in Fig. 4(a), where the tracked FPs are denoted by symbol '+'. The FPs were extracted by the Harris interest operator [21]. The observation matrix of the video frames is presented in Fig. 4(b), where red points are the observed entries. The proposed algorithm is verified for noise levels from zero to 2 [pixel].

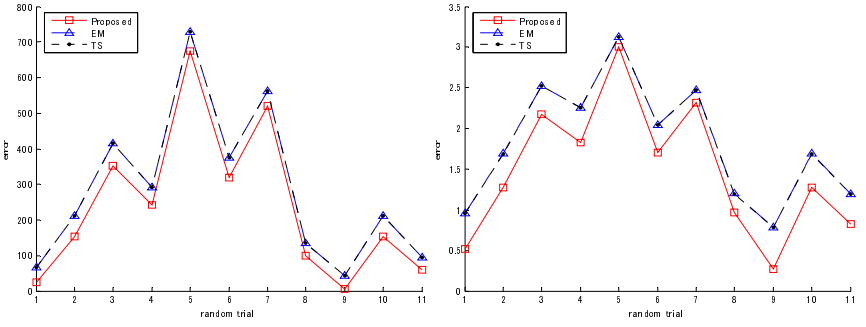


(a)



(b)

Fig. 4. Real video seuqnece. (a) 3rd, 6th , and 9th frame of 'dinosaur' sequence with the tracked FPs. (b) observation matrix (36x4983).



(a) (b)

Fig. 5. *RMS* errors. (a) 2D reprojection errors (2D-*RMS* errors). (b) 3D reconstruction errors (3D-*pRMS* errors).

Fig. 5(a)-(b) illustrates the results of the estimated missing data. The proposed method leads to results of greater accuracy for the 2D-*RMS* errors and 3D-*pRMS* errors in various levels of noise. Therefore, it can be confirmed that the proposed method provides more satisfied results for real sequences.

6 Conclusions

In order to handle the problem of missing data, a new geometrical approach is presented by minimizing the influence of noise. The main idea is to estimate missing data by the geometrical correlations between 2D image plane and 3D error space. The achievements of the proposed system are presented using experimental results using real sequence. In the results, it can be confirmed that the proposed geometrical fitting algorithm can provide more accurate estimation results than the EM and TS algorithms' in various levels of noise. The abilities of the proposed system are as follows: i) converge to accurate estimation by calculating a wide ranges of noise levels, ii) handle the orientation and distance of a missing data by geometrical correlations.

In the near future, the system will be extended to estimate noise level within a recovered 3D shape. This estimation algorithm will support the creation of 3D video material from original 2D footage with self-occluding objects. Furthermore, the proposed method will assist 3D content creation by advancing 3D TV significantly and increasing its attractiveness.

References

1. C. Fehn, P. Kauff, M. Op de Beeck, F. Ernst, W. I.Jsselsteijn, M. Polleys, L. Van Gool, E. Ofek and I. Sexton : An Evolutionary and Optimized Approach on 3D-TV. In Proc. IBC02 (2002) 357-365
2. M. op de Beeck, P. Wilinski, C. Fehn, P. Kauff, W. I.Jsselsteijn, M. Polleys, L. Van Gool, E. Ofek and I. Sexton : Towards an Optimized 3D Broadcast Chain. In Proc. SPIE IT-Com02 (2002) 357-365
3. Tomasi and T. Kanade : Shape and motion from image streams under orthography: A factorization method. Int. J. of Computer Vision, vol.9-2 (1992) 137-154
4. Sturm, P., Triggs, B.: A factorization based algorithm for multi-image projective structure and motion. In Proc. ECCV, vol. 2, Cambridge, UK (1996) 709-720
5. Poelman, C.J., Kanade, T. : A paraperspective factorization method for shape and motion recovery. IEEE Trans. on PAMI, vol.19-3 (1997) 206-218
6. Morita, T., Kanade, T. : A sequential factorization method for recovering shape and motion from image streams. IEEE Trans. on PAMI, vol.19-8 (1997) 858-867
7. Aguiar, P.M.Q., Moura, J.M.F. : Rank 1 Weighted Factorization for 3D Structure Recovery: Algorithms and Performance Analysis. IEEE Trans. on PAMI, vol.25-9 (2003) 1134-1149
8. Irani, M., Anandan, P. : Factorization with uncertainty. In Proc.ECCV, Dublin, Ireland (2000) 539-553
9. Shapiro, L., : Affine Analysis of Image Sequences. Cambridge University Press (1995)
10. Quan, L., Kanade, T. : A factorization method for affine structure from line correspondences. In Proc. IEEE CVPR, San Francisco CA, USA (1996) 803-808
11. Morris, D., Kanade, T. : A unified factorization algorithm for points, line segments and planes with uncertainty models. In Proc. IEEE ICCV (1998) 696-702
12. Aguiar, P.M.Q., Moura, J.M.F. : Three-dimensional modeling from two dimensional video. IEEE Trans. on Image Processing, vo.10-10 (2001)1544-1551
13. D. Jacobs : Linear fitting with missing data for structure-from-motion. In Proc. CVIU, vol.82 (2001) 57-81

14. Rui F. C. Guerreiro, Pedro M. Q. Aguiar : Estimation of Rank Deficient Matrices from Partial Observations: Two-Step Iterative Algorithms. In Proc. EMMCVPR (2003) 450-466
15. Zhaohui Sun, V. Ramesh, and A. Murat Tekalp : Error characterization of the factorization method. Computer Vision and Image Understanding, vol.82 (2001) 110-137
16. K. Daniilidis and M. Spetsakis : Understanding noise sensitivity in structure from motion, Visual Navigation (1996) 61–88
17. D. D. Morris, K. Kanatani, and T. Kanade : Uncertainty modeling for optimal structure from motion. In Vision Algorithms: Theory and Practice, Lecture Notes in Computer Science, Springer-Verlag, Berlin/New York, vol.1883 (2000) 200–217
18. R. Szeliski and S. B. Kang : Shape ambiguities in structure-from-motion. IEEE Trans. on PAMI. vol.19 (1997) 506–512
19. R. Szeliski, and S.B. Kang : Shape Ambiguities in Structure From Motion. Technical Report 96/1, Digital Equipment Corporation, Cambridge Research Lab, Cambridge, Mass. (1996)
20. Fitzgibbon, G. Cross, and A. Zisserman : Automatic 3D Model Construction for Turn-Table Sequences, 3D Structure from Multiple Images of Large-Scale environments. In Proc. SMILE Workshop on Structure from Multiple Images in Large Scale Environments, Lecture Notes in Computer Science, vol.1506 (1998) 155-170
21. C. J. Harris and M. Stephens : A Combined Corner and Edge Detector. In Proc. Alvey Vision Conf. (1998) 147-151

A Flame Detection Algorithm Based on Video Multi-feature Fusion

Jinhua Zhang, Jian Zhuang, Haifeng Du, Sun'an Wang, and Xiaohu Li

Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University, 710049 Xi'an, China
jjshua@mail.xjtu.edu.cn

Abstract. The physical flame characteristics in the burning process are analyzed in detail. Combined with static and dynamic flame features, a flame detection algorithm based on multi-feature fusion is designed to quickly recognize fire flame using the theory of the degree of belief. The experiments show that the processing rate of the algorithm is as high as 25 frames/s without any manual adjustments. In complex situations, the algorithm remains with stronger anti-jamming ability and higher accuracy, for example, the suspected probability of street lamps and automotive lamps is lower than 10%. From comparing results, the video-based detection algorithm has the good capabilities.

1 Introduction

Fire detection is a very important issue because it is closely related to people's safety and property. Fire detection techniques have been explored by many investigators and video-based approach is becoming more and more interesting. Flame is an important characteristic of flaming combustion and plays an important role in fire detection. Now, most fire detection methods based on video image use flame as a primary criterion to recognize fire. These methods have many distinct advantages, such as discovering fire quickly and finding fire location directly. Hence, they have attracted much attention in the field of automatic fire surveillance, especially in the field of fire detection in large space.

Many studies have been conducted on fire flame detection algorithms based on video. And most flame detection algorithms mainly focus on the analysis of flame color and flame area. These algorithms commonly extract flame combustion regions according to setting threshold in the color space, and eliminate surrounding disturbance by area criterion. They have the advantages of rapid processing rate and good real time ability. But, they can't work well under complex situations because of the disturbance from the surrounding environment.

In some works, the shape characteristics are incorporated into flame detection criterion. Liu et al. [1] presented spectral, spatial and temporal models of fire regions in visual image sequences. They suggested that the shape of a fire region was represented in terms of the spatial frequency content of the region contour using its Fourier coefficients and the temporal changes in these coefficients were used as the temporal signatures of the fire region. Results given in the literature showed that their method was capable of detecting fire reliably. But the method proposed by Liu has a strict

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 784 – 792, 2006.

requirement for the environments. And the method can not detect flame in some situations, such as low burning power and relatively steady burning. Li et al. [2] used the first-order moments of consecutive flames as a criterion to detect flame. Their experiments showed that the first-order moments of all kinds of flames were not affected by the different burning materials and the viewing distance. Moreover, for the first-order moments, the average number of the changing pixels of flame was greater than that of flame-like disturbances. However, under the complex illumination condition (for example, with the disturbance due to moving vehicle lamps), the first-order moments of non-fire regions have great change, which results in the un-successful fire flame detection by their method.

In order to inform user of fire alarm as early as possible, the practical image based fire monitoring system must work in real time and must be applied in more general environments instead of being restricted on some specific situations. Accordingly, a better flame detection algorithm should have stronger anti-jamming ability and rapid processing rate in order to ensure the accuracy, robustness and real-time ability of fire detection. In this paper, the static and dynamic features of fire flame in burning process are analyzed and a video-based flame detection algorithm is proposed. The algorithm synthesizes multi-features of fire flame to detect flame, and works out the occurrence probability of fire regions according to the probability model of the flame detection. The experiments show that our algorithm can work well in the complex situations and detect fire accurately and quickly at the rate of 25 frames a second.

The organization of this paper is as follows. In section 2, we discuss the physical flame characteristics in burning process, including the static and dynamic features. Section 3 proposes the probability model of flame detection and flame detection algorithm. Section 4 shows the experiment results. Finally, the conclusions are given in section 5.

2 Physical Flame Characteristic

2.1 Static Flame Feature

Flame can be divided into two categories: static flame and diffusion flame. During combustion, fuels are in contact with air. So the burning process depends on the mutual diffusion rate of fuel and air. The flame producing in this process is called diffusion flame. Diffusion flame emits light and releases heat in burning process. Generally, the temperature of flame gas is as high as over 1400K. Furthermore, the shape of diffusion flame keeps changing and exhibits a lot of shape features, such as burning height, flame core, flame spire points. However, flame shape features depend on the environmental factors, for instance airflow and burning material [3]. In video images, the pixels in fire region display reddish colors and fire regions are distinctly brighter than other regions under ideal lighting conditions [4, 5]. Fire region exhibits a structure of nested rings of closed contours because of the gradient difference between flame core and periphery, and moreover, one or more spire points exist in the contour [6]. They represent the characteristics of light and heat emission and the shape features of flame, respectively.

Suppose that in *n*th frame of the video sequence, after image segmentation, the *m*th region is a flame region. Then the pixel set of the region is

$$\mathrm{Cl}_n^m = \left\{ \mathrm{P}_{ij} \mid f(\mathrm{P}_{ij}) > s_r, \mathrm{P}_{ij} \subseteq \mathrm{MAP} \right\}, \tag{1}$$

where P_{ij} is a pixel with a triple (r, g, b); r, g, and b are the intensities of the R, G, and B components respectively in RGB color model and P_{ij} is adjacent with other pixels in the same set, MAP is a set of M*N digital images, $f(\mathrm{P}_{ij})$ is the estimate function of the flame chroma and luminance, s_r denotes a threshold.

Many researchers have studied the eigen function of chroma and luminance and have presented a lot of eigen functions in different styles. However, by single eigen function, it is difficult to distinguish between fire flame area and flame-like area in the complex situation. As shown in Fig.1, the left is an original flame image in the street at night. In the original image, there are many regions in which pixels have the same values of r, g, and b of the RGB model with pixels in the flame region, such as vehicle lamp region, street lamp region and the reflection region of the ground. The right is the segmentation result obtained by threshold segmentation algorithm. The result shows that the disturbance area, street lamp region and vehicle lamp region can't be eliminated only using simple chroma function or luminance function. Under such conditions, the detection algorithms are difficult to recognize fire flame area correctly. It is obvious that the methods based on simple chroma fuction or luminance function have strict requirement for detection situations.

Fig. 1. Street flame image: Left: The original image; Right: The processed image by threshold segmentation

In video image, the closed contour set of flame area $\mathrm{Cl}^m{}_n$ is defined by

$$\mathrm{Contour}(\mathrm{Cl}_n^m) = \{\mathrm{P}_0, \cdots \mathrm{P}_k, \mathrm{P}_0\}, \tag{2}$$

where P_k is a pixel in the edge of the area. When edge trembling exists in the contour, it must be satisfied with the following relation

$$\theta(\overrightarrow{\mathrm{P}_{k-1}\mathrm{P}_k}, \overrightarrow{\mathrm{P}_{k+1}\mathrm{P}_k}) \le \theta_0, \tag{3}$$

Where θ() is the angle function of two vectors and θ_0 is a angle threshold.

According to combining static flame features, closed contours and edge trembling together, flame detection algorithm can eliminate most of the surrounding disturbances. Nevertheless, false recognition would happen in some situations. For example, edge trembling of vehicle lamps and background would result in unsuccessful flame detection only by static flame features. Therefore, in order to improve the capability of removing spurious flame-like regions, flame dynamic features are needed to integrate the detection algorithm.

2.2 Dynamic Flame Features

Diffusion flame is flickering in burning process, and the flicker phenomenon is related to tangent force between fuel jets and ambient air. Flicker frequency is low and normally from several hertz to twenties [3]. Image sampling frequency is 25HZ, namely 25 frames per second in video sequence. Obviously, flame flicker can not be sampled without any distortion at the sampling frequency. Consequently, it is difficult to obtain feature spectrum of flame flicker directly by video information. Whereas we find that the height of flame is changeable due to flame flicker, the height change pattern is directly related to flame flicker frequency and the change pattern differs from jamming sources in our practical operation process. So, we use the flame height change as a dynamic eigen-factor to detect flame.

Let us suppose that H_m is the height sequence set of fire-like area CI^m extracting from video image sequence. Then the elements of the set can be written as

$$h_i^m = \text{Height}(\text{Cl}_i^m) . \tag{4}$$

Define A_m as the set of the coefficients of the Discrete Fourier Transform (DFT) of the height sequence H_m. Then the element of A_m is described as

$$a_k^m = \text{DFT}(\text{H}^m) = \frac{1}{n}\sum_{i=1}^{n} h_i^m \exp(-j\frac{2\pi}{n}ik) . \tag{5}$$

Here the height change eigen function is

$$\text{f}_d(\text{A}^m) = \sum_{i=2}^{\frac{l}{2}} \frac{a_k^m \cdot \overline{a_k^m}}{l/2-1} , \tag{6}$$

where l is the length of the Fourier transform, $a^m{}_k$ is a coefficient of the DFT. The larger the value of $f(A_m)$ is, the more spectrum components are.

Three video sequences have been selected to test the performance of our algorithm. They are vehicle lamp video sequence, street lamp video sequence, and fire flame video sequence, respectively. And four sample frames from each sequence are shown in Fig.2. The first sequence is called "Vehicle", which is 400 frames length showing a vehicle at night. The sequence contains a vehicle lamp that moves from far to near. And the sample frames are shown in the top row of Fig.2. The second sequence is called "Street", which are 400 frames length showing the street lamp in the night. And the sample frames are shown in the second row. The third sequence "Flame" is also 400 frames length. And the sample frames are shown in the bottom row of Fig.2.

Fig. 2. Video image sequences. Top row: Vehicle video sequence; Middle row: Street video sequence; Bottom row: Flame video sequence.

The proposed method was used to detect flame after image segmentation. For the three video sequences, the spectrograms of the height change of the detected object

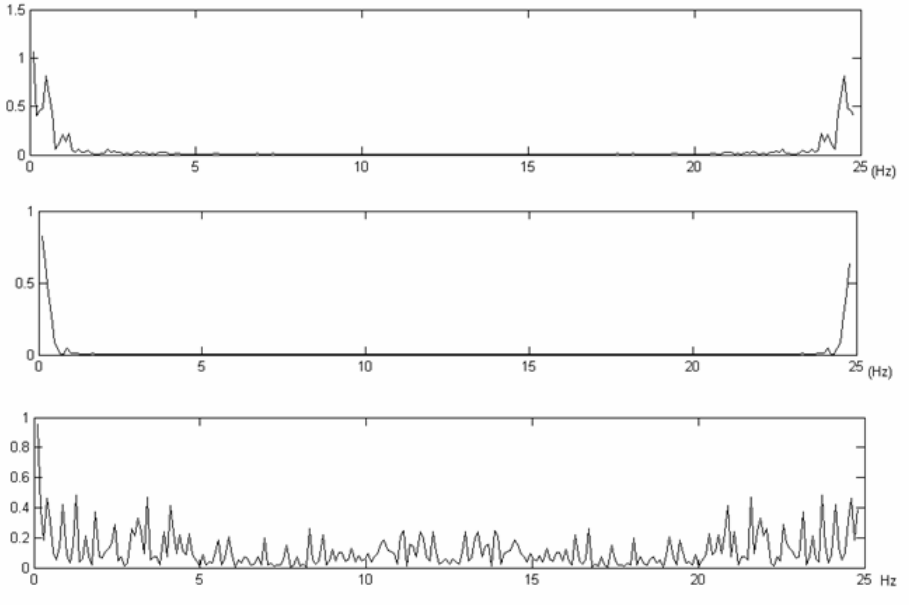


Fig. 3. Spectrogram of the height change. Top: Vehicle spectrogram; Middle: Street spectrogram; Bottom: Flame spectrogram.

are shown in Fig.3. The length of the Fourier transform was 256 points. The top of Fig.3 is the spectrogram of the object height change in "Vehicle"; the middle is the spectrogram of the object height change in "Street", and the bottom is in "Flame".

Comparing the three spectrograms, we found that the spectrum of the “Flame” video sequences contained very abundant spectrum components and two others had few spectrum components. Thus, this feature was used as a primary criterion to detect flame region. In order to reflect the change of the signal spectrum, more sample signals may be needed. The sample signals contain many dynamic random disturbances with more spectrum components and disturbances occur intermittently and remain shortly, such as the lamp of moving vehicle. But the time to detect flame combustion could be longer than the duration of disturbance. So the probability model of fire alarm is established and the uncertainty of recognizing flame is decreased by increasing sample video sequence.

3 Probability Model and Flame Detection Algorithm

Suppose that H is a hypothesis and E is an item of evidence. Thus CF(H|E), the certainty factor of hypothesis H given evidence E, is used as the uncertainty of measurement and denotes the modified quality of the subjective degree of belief in the certainty factor model. At the same time CF(H|E) reflects that people increase or decrease the belief of uncertain knowledge. Based on the certainty factor model and the flame characteristics of video images, the probability model of flame detection is established as following

$$CF_n^m = \begin{cases} k\,\mathrm{f}_d(\mathrm{A}^m)\sum_{i=1}^{n}\mathrm{f}_1(\mathrm{Cl}_n^m) & 0 < CF_{n-1}^m < 1 \\ 1 & CF_{n-1}^m \geq 1 \\ 0 & CF_{n-1}^m \leq 0 \end{cases}, \tag{7}$$

$$\mathrm{f}_1(\mathrm{Cl}_m^n) = \begin{cases} a & \mathrm{Cl}_m^n \text{ area exists with edge trembling} \\ -b & \mathrm{Cl}_m^n \text{ area don't exist or without edge trembling} \end{cases},$$

where k is an adjusting coefficient, the values of a and b are defined by the following rule: if allowing false alarms instead of neglect, then a is a great number and $a>b$; if no false alarms, then a is a little number and $a<b$.

Accordingly, based on the analysis of the physical flame characteristics and the probability model of the flame detection, a new flame detection algorithm based on multi-feature fusion (FDAMF) is proposed. The algorithm is described as follows.

1. Judge whether user ends the process. If user ends the process, then jump into step 7. Otherwise capture video images and put into the set MAP.
2. Segment images in the set MAP using image segmentation algorithm based on Thermodynamics [7]. Extract closed regions $\mathrm{Cl}^m{}_n$ according to equation (1) and calculate the coefficients of the Fourier transform of the height change (The change

length is 256 points. If the length of data is less than 256, the length of data is extended to 256 points with zero.
3. Calculate the probability CF^m_n of flame occurrence in each region by the formula (7).
4. Judge whether $CF^m_n>1$. If $CF^m_n>1$, then alarm. Or else show the maximal suspected probability in the suspected region set.
5. Judge whether $CF^m_n = 0$. If $CF^m_n = 0$, then discard the suspected region from the suspected region set.
6. Jump into step (1).
7. End the process.

4 Experiments and Analysis

The experiments were carried out using C++ on a PC with PIV3.0G CPU and 1GB primary memory. The size of video sequence image is 300×300 pixels. Four sample frames from the processed results of three test video sequences are shown in Fig.4.

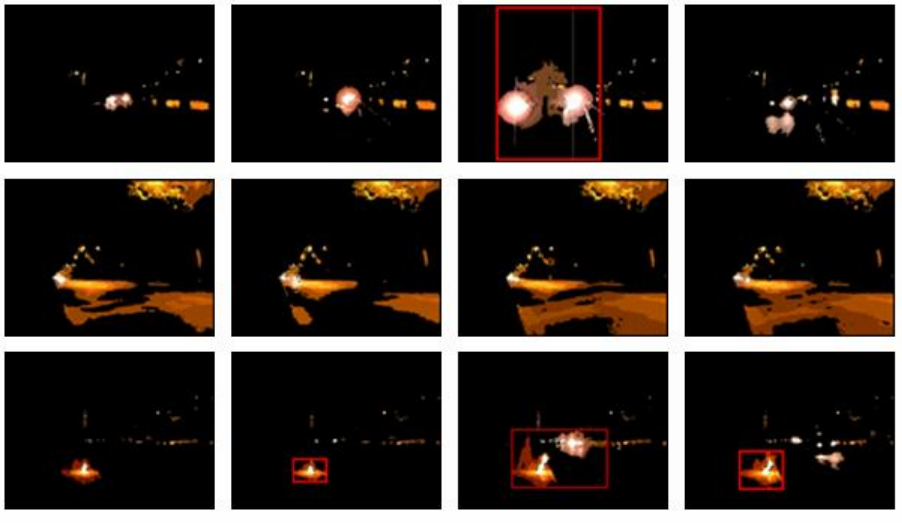


Fig. 4. The processed results of three video sequences. Top row: The result video sequence of vehicle; The second row: The result video sequence of street; Bottom row: The result video sequence of flame.

The top row of Fig.4 is the processed result of vehicle video sequence. The second row is the processed result of street video sequence. The bottom row is the result of flame video sequence. In the processing procedure, if the suspected probability of the region is greater than 10%, then the region is marked with a red rectangle. As shown in the first row and the third column of Fig.4, some moving vehicle lamp regions marked with a red rectangle are suspected of flame areas in “Vehicle” sequence. But the moving vehicle lamp remains in the video sequence for a while, the suspected

probability reduces to zero quickly after the vehicle lamp disappears according to the formula (7). As shown in the middle row of Fig.4, the static features of the reflection of the ground and the leaves are similar to those of flame areas, but their dynamic features are different to those of flame in the street lamp video sequence. Hence, the probability of all closed areas is less than 10% and no red rectangle occurs in the video sequence. In the flame video sequence, shown in the last row of Fig.4, our method recognizes the flame area quickly. At the same time the suspected probability increases to 100%, then the system gives a fire alarm. Because the position of flame is marked by a simple maximal rectangle, the vehicle lamp could be included in the flame region when they are closed to the fir flame. The experiments demonstrate that our flame detection algorithm can locate the position of flame automatically, has a rapid processing rate and the time processing each frame is less than 40ms. The algorithm still has strong anti-jamming ability and can be used in practical environment.

In order to verify the performance of our algorithm, we have done three comparing experiments that correspond to the three algorithms: the flame detection algorithm by the first-moments (FDAFM), the flame detection algorithm by the Fourier coefficients of the flame shape (FDAFC) and our method mentioned above. Under indoor and outdoor environments, the flame was extracted from the high power combustion and the low power combustion using the three algorithms. The combustion experiments, with light source disturbance, were repeated 100 times in different environments. The duration of each experiment was 10 minutes. The statistical results of the correct alarm rate and the false alarm rate of the three methods are given in Table 1.

Table1 shows that the flame detection algorithm based on video multi-feature fusion is better than others. This method can recognize flame correctly and has no false

Table 1. The capabilities of three algorithms. The first colum denotes algorithm name; The second column is the correct alarm rate in the indoor high power combustion experiment (CARIHPC); The third is the false alarm rate in the indoor hige power combustion experiment (FARIHPC); The fourth is the correct alarm rate in the indoor low power combustion expriment (CARILPC); The fifth is the false alarm rate in the indoor low power combustion experiment (FARILPC); The sixth is the correct alarm rate in the outdoor high power combustion experiment (CAROHPC); The seventh is the false alarm rate in the outdoor high power combustion experiment (FAROHPC); The eighth is the correct alarm rate in the outdoor low power combustion experiment (CAROLPC); The last is the false alarm rate in the outdoor low power combustion experiment (FAROLPC).

Algo-rithm	CARI HPC	FARI HPC	CARI LPC	FARI LPC	CAR OHPC	FARO HPC	CAR OLPC	FARO LPC
FDAF M	87%	12%	67%	12%	56%	23%	34%	35%
FDAF C	95%	3%	56%	6%	90%	8%	46%	17%
FDAM F	100%	0%	95%	0%	100%	0%	93%	0%

alarm in high power combustion and low power combustion, no matter under indoor situation or outdoor situation. Hence, our method can work well in practical use.

5 Conclusions

In the paper, we analyzed the static and dynamic features of fire flame and designed a flame detection method based on multi-features fusion. Then we proposed the probability model of flame occurrence based on the theory of the degree of belief and used the suspected probability to represent the probability of fire flame occurrence in video image. The experiment results show that the detection algorithm processes image quickly at rate of 25 frames a second and detects flame with high accuracy. Moreover, our method has strong anti-jamming ability and can be applied to practical complex environment. The program code and the video sequences, in the paper, can be downloaded from http://202.117.58.58/files/FlameAlarm_setup.exe.

However, the ability of our method can be improved further. For instance, we used a rectangle to mark the flame, which was easy to treat disturbance as flame and resulted in deviation of the location of fire flame occurrence. On the other hand, how to remove the jamming source under strong light in the day is our future work.

Acknowledgement

This work is supported by the National Natural Science Foundation of China under Grant No.50505034.

References

1. Liu, C.B., Ahuja, N.: Vision Based Fire Detection. Proceedings of the 17th International Conference on Pattern Recognition. IEEE Press, Cambridge UK (2004) 134-137
2. Li, J., Fong, N.K., Chow, W.K. et al.: The Motion Analysis of Fire Video Images Based on Momemt Features and Flicker Frequency. Journal of Marine Science and Application, Vol 3. (2004) 81-86.
3. Wang, F.: Flames, their Structure, Radiation and Temperature. Science Press, Beijing China (1994) 84-112 (in Chinese)
4. Horng, W.B., Peng, J.W., Chen, C.Y.: A New Image-Based Real-Time Flame Detection Method Using Color Analysis. Proceedings of the 2005 IEEE International Conference on Networking, Sensing and Control. IEEE Press, Tucson Arizona USA (2005) 100-105
5. Chen, T.H., Wu, P.H., Chiou, Y.C.: An Early Fire-Detection Method Based on Image Processing. Proceedings of the 2004 International Conference on Image Processing. IEEE Press, Singapore (2004) 1707–1710
6. Wu, L.B., Song, W.G., Lu, J.C.: A New Fire Recognition Method in Image Fire Monitoring. Fire Safety Science. Vol 2. (1997) 61-65
7. Zhuang, J., Yu, Q., Wang, S.A.: The Study of the Image Segmentation Algorithm Based on Thermodynamics. Journal of Xi'an Jiaotong University. Vol 39. Xi'an (2005) 507-510 (in Chinese)

An Accelerated Algorithm of Constructing General High-Order Mandelbrot and Julia Sets

Chong Fu[1], Hui-yan Jiang[2], Xiu-shuang Yi[2], and Zhen-chuan Zhang[1]

[1] School of Information Science and Engineering, Northeastern University, 110004 Shenyang, China
fu_chong@sohu.com
[2] Computing Center, Northeastern University, 110004 Shenyang, China

Abstract. A series of high order Mandelbrot and Julia fractal images of complex map $f: z \leftarrow z^n + c (n > 2, n \in N)$ are constructed by using escape time algorithm. Topology invariance and the relation between period bud distribution and mapping orders are found through lots of computer mathematic experiments. The escape time algorithm is improved on basis of the rotation symmetric property of Manderlbrot and Julia sets, and then a rotation escape time algorithm is established. The high-order Mandelbrot and Julia sets fractal images are generated by using Web oriented Java Applet. Therefore the complexity of computer simulation under complex conditions is greatly reduced. The application of Java technology provides an Internet-based distribution platform for chaos-fractal theory study.

1 Introduction

Mandelbrot and Julia sets has been regarded as one of the most complicated collections in mathematical field since American mathematician Benoit B.Mandelbrot of IBM Corporation first constructed the images of Mandelbrot set in 1980 [1]. Scientists have discovered many inner rules and fractal properties of Mandelbrot and Julia sets through 20 years research [2]. Up to now, many questions in mathematics and computer graphical fields have not been solved yet and people will continue to have a research on Mandelbrot and Julia sets for the valuable reason. In this paper, the escape time algorithm on constructing general high-order Mandelbrot and Julia sets are improved through lots of computer mathematics experiments. The plotting speed is greatly improved. The topology invariance property of general high-order Mandelbrot and Julia sets and the embedded rule of period bud in fractal space are also discussed.

2 Basic Theories for Mandelbrot and Julia Sets

Definition 1. The collection of complex parameter c that makes the sequence $c, c^n + c, (c^n + c)^n + c, \cdots$ constructed by high-order complex map $f: z \leftarrow z^n + c$ $(n > 2, n \in N)$ bounded is called a kind of general Mandelbrot set, or *M* set for short, namely

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 793 – 796, 2006.

$$M = \{c \in C \mid c, c^n + c, (c^n + c)^n + c, \ldots \text{is bounded}\}. \quad (1)$$

Definition 2. The collection of complex parameter z that makes the sequence $z, z^n + c_0, (z^n + c_0)^n + c_0, \cdots$ constructed by high-order complex map $f : z \leftarrow z^n + c$ $(n > 2, n \in N)$ bounded is called a kind of general Julia set, or *J* set for short, namely

$$J = \{z \in C \mid z, z^n + c_0, (z^n + c_0)^n + c_0, \ldots \text{is bounded}\}. \quad (2)$$

Definition 3. $\forall x_0 \in M$, If p is the minimum integer for $f^p(x_0) = x_0$, then x_0 is called the p periodic point of general *M* set. This different single connective domain that contains same period points is called the period bud of general *M* set.

Definition 4. Let x_0 be a periodic point of period p , if complex derivative $(f^p)'(x_0) = 0$, then x_0 is called the super attractor periodic point of f .

To a non-linear dynamical system, super attractor point is the steadiest and the stable condition of period point can be extended as

$$|q| < 1, \quad (3)$$

while $|q| = 1$ (called neutral points) corresponds to the boundary of steady domain.

Fractal space is a topology space that defined by Hausdorff metric, whereas fractal set is a topology embedded invariance set constructed by compression map in fractal space.

Definition 5. The δ-parallel object of *A* is a close set construct by the points whose distance to *A* is not greater than δ, namely

$$A_\delta = \{x \in X \mid \rho(x, A) \le \delta\}, \quad (4)$$

then

$$h_\rho(A, B) = \inf\{\delta \mid A \in B_\delta \text{ and } B \in A_\delta\} \quad (5)$$

is called the Hausdorff distance from A to B.

Theorem 1. Let (X, ρ) be self-contained metric space, then $(H(X), h_\rho)$ is self-contained space, called fractal space, and if $\{A_n \in H(X)\}_{n=1}^{\infty}$ is a Cauchy sequence, then $A = \lim_{n \to \infty} A_n$ can be presented as

$$A = \{x \in X \mid \exists \textit{Cauchy sequence}\{x_n \in A_n\} \text{ convergence to } x\}. \quad (6)$$

Theorem 2. Let $w : X \to X$ be a compression map with compression ratio *c* in self-contained metric space (X, ρ) , then $W : H(X) \to H(X)$ defined by the following equation is a compression map with compression ratio *c* on $(H(X), h_\rho)$:

$$w(B) = \{w(x) \mid x \in B\}. \quad (7)$$

3 Rotation Escape Time Algorithm

Escape time algorithm is a common and effective method for plotting *M* and *J* sets [3-5], but for high-order *M* and *J* sets, when the order grows, the calculation complexity grows exponentially. For example, when $m=10$, for a size of 640×480 image domain, it takes approximate 1 billion times calculation to iterate 30 times. Effectively reducing the calculating time and storage space complexity is an important guarantee for the study of high-order questions. The symmetric relationship between order and period bud and origin and axes has been found by constructing a series of general high-order *M* and *J* sets. The escape time algorithm is improved by using this property, namely rotation escape time algorithm. The following theorem is the theoretic basis that the entire rotation symmetric domain in *M* and *J* fractal sets has the same escape time.

Theorem 3. For general *M* set constructed by high-order complex map $f: z \leftarrow z^n + c \ \ (n \geq 2, n \in N)$,

$$| f^k(c) | = | f^k(ce^{i\frac{2\pi j}{n-1}}) |, k = 1,2,\cdots,N; j = 0,1,2,\cdots,n-2. \tag{8}$$

The rotation escape time algorithm is constructed as follows:

(1) Select a maximum iteration step *N* that computer calculates;
(2) Select an escape radius $R \geq 2$;
(3) Define escape time function as:

$$T(x) = \begin{cases} k & |f^k(x)| \geq R, \ |f^i(x)| < R \\ & i = 1,2,\cdots,k \ \ \mathrm{k} \leq \mathrm{N} \\ 0 & |f^i(x)| < R, \ i = 1,2,\cdots,N \end{cases}. \tag{9}$$

(4) For integer $n \geq 2$, select a sector domain $0 \leq \theta \leq \frac{2\pi}{n-1}$ (θ is extent angle), and do iteration calculation to each point *x* in this sector domain. If it is an attractor, the escape time function $T(x) = 0$, then rotating an angle of $\frac{2\pi j}{n-1}$ $(j = 1,2,\cdots,n-2)$ in turn, the points obtained from the rotation are also attractors.

(5) Color each point according to the escape times of escape time function.

4 Experimental Results and Conclusions

The rotation escape time algorithm greatly reduces the time and space complexity of calculating with the plotting time of *M* and *J* fractal sets approaching $1/(n-1)$ of the original escape time algorithm. The *M* and *J* sets fractal images plotted by rotation escape time algorithm are shown in Fig. 1 and 2. The program is implemented by using Web oriented Java Applet, which provides a distributed and open platform for

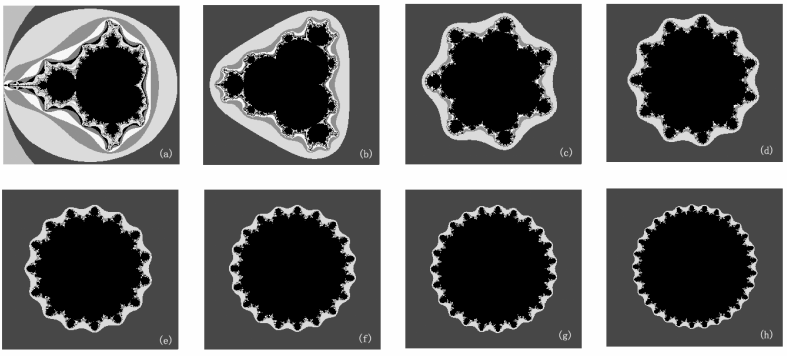
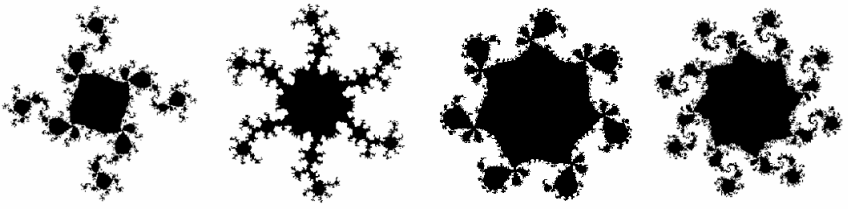


Fig. 1. General high order *M* set fractal images
(a) m=2 ; (b) m=4 ; (c) m=8 ; (d) m=12 ; (e) m=16 ; (f) m=20 ; (g) m=24 ; (h) m=30



(a) (b) (c) (d)

Fig. 2. General high order *J* set fractal images
(a) m=5 ; (b) m=7 ; (c) m=8 ; (d) m=9

chaos-fractal theory study. The paint method and Graphics object in Applet class provide a uniform interface for drawing graphics in a Web page.

General *M* set is not topologically homeomorphic with *M* set once $n > 2$. The high-order *M* and *J* fractal images are all rotation symmetric and have total $n-1$ sector symmetric domains. The number of period 2 buds of general high-order *M* sets is *n*-1, and the buds distribute equably at the circumference of period 1 bud.

References

1. Mandelbort, B. B.: The Fractal Geometry of Nature. San Francisco: Freeman (1982) 1-15
2. Peitgen, H. O., Saupe, D.: The science of fractal images. Berlin: Springer-Verlag (1998) 23-27
3. Rojas, R.: A tutorial on efficient computer graphic representation of the Mandelbrot set. Computers and Graphics, Vol.15 (1991), 91-100
4. Yan, D. J., Liu, X. D., Zhu, W. Y.: An investigation of Mandelbrot set and Julia sets generated from a general complex cubic iteration. Fractal, Vol.7 (1999), 433-437
5. Liu, X. D., Zhu, W. Y.: Composed accelerated escape time algorithm to construct the general Mandelbrot sets. Fractal, Vol.9 (2001), 149-153

A Novel Approach Using Edge Detection Information for Texture Based Image Retrieval

Jing Zhang[1] and Seok-Wun Ha[2]

[1] Dept. of computer science,GyeongSang National University, Korea
zhj2002cn@hotmail.com
[2] Dept. of computer science, RICIC&ERI,GyeongSang National University, Korea
swha@gsnu.ac.kr

Abstract. Most texture-based image retrieval system just consider an original image of coarseness, contrast and roughness, actually there are many texture information in the edge image. In this paper, a method combining both edge information and gray level co-occurrence matrix properties is proposed to improve the retrieval performance. The proposed method gives encouraging results when comparing its retrieval performance to that of the Yao's method, in the same image database.

1 Introduction

Texture, the variations of shading or color, is an important feature in images. Texture analysis has long been an active area in computer vision. While the human ability to distinguish different textures is obvious, the automated description and recognition of these patterns have proven to be quite complex. There are many algorithms expounded how to extract texture feature efficiently and accurately in the field of image retrieval. W. S. Zhao[1] used the angle and gradient magnitude to represent the texture directionality. N.V. Shirahatti and K. Barnard[2] evaluated query-result pairs for both query by image example and query by text. M.Partio, B.Cramariuc and M.Gabbouj[3] proposed a combination of ordinal measures and co-occurrence matrices. In Yao's[5] method, a smoothed image is obtained by using Sobel operator, then construct a gray level co-occurrence matrix(GLCM) with the correspond pixels in both of these two images and a vector composed with the set of matrix statistical features is regarded as the final characterization of the image.

In this work, image retrieval based on texture is studied, and a novel approach to retrieve images proposed for texture retrieval, the proposed method used a composition of edge detection information and GLCM properties oriented to four directions to extract texture feature. Retrieval performance of the proposed method is evaluated using a set of well known Brodatz textures[8], which consists of 10 classes, each class contain 30 similar images with size 128*128.

The remainder of the paper is organized as following. Section 2 describes the compass edge detector and GLCM. Retrieving images and performance evaluation are presented in section 3, and the paper is concluded in section 4.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 797 – 800, 2006.

2 Proposed Method

2.1 Compass Edge Detector

The edges of an image hold one of the texture information in that image. The edges tell where objects are, their shape and size, and something about their texture[4]. When using compass edge detection the image is convolved with a set of convolution kernels, each of which is sensitive to edges in a different orientation. In order to combine with four directions gray level co-occurrence matrices we only use four kernels of eight: 0°, 45°, 90° and 135°. We select Prewitt operator because it is not susceptible to noise. Figure 1 (A) shows the directions of edge detection.

2.2 Gray Level Co-occurrence Matrix

The gray level co-occurrence matrix defined by Haralick[7] can reveal certain properties about the spatial distribution of the gray levels in the texture image. It denotes how often a pixel with the intensity value i occurs in a specific spatial relationship to a pixel with the value j. Each element (i,j) is simply the sum of the number of times that the pixel with value i occurred in the specified spatial relationship to a pixel with value j. By default, the spatial relationship is defined as the pixel of interest and the pixel to its horizontally adjacent. In this work, four spatial relationships were specified: 0^0, 45^0, 90^0 and 135^0, Figure 1（B） shows it . In order to avoid large co-occurrence matrices, the number of intensity values in grayscale image is used scaling to reduce from 256 to 8.

In order to estimate the similarity between different gray level co-occurrence matrices, many statistical features extracted from them were proposed[6]. To reduce the computational complexity in this study we use the four of them: energy, entropy, contrast and homogeneity.

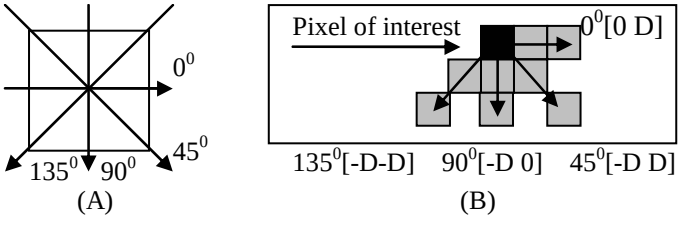


Fig. 1. Directions for extracting texture feature. (A) Edge detection directions; (B) Co-occurrence matrix directions.

2.3 Combination of Edge Information and GLCM Properties

The output of edge detectors can form continuous boundaries of the interest regions that can be used for extracting texture feature in content based image retrieval(CBIR) system. So firstly the detection of edge is done for query image using only four among eight direction kernels of Prewitt edge detector. When using compass edge detection the image is convolved with a set of convolution kernels. For each pixel of

the query image the local edge gradient magnitude is estimated with the response of four kernels at the pixel location, and we can respectively obtain 4 direction edge images. Through edge detecting, 4 texture images are formed. In these texture edge images there are salience spatial relationships between different gray level pixels which we interest.

On the other hand, the GLCM can reveal certain properties about the spatial distribution of the gray levels in the texture image, so we respectively construct same direction GLCM from corresponding direction texture edge image, and then compute all gray level co-occurrence matrices statistical features: contrast, entropy, energy, and homogeneity. Figure 2 shows the process of combination.

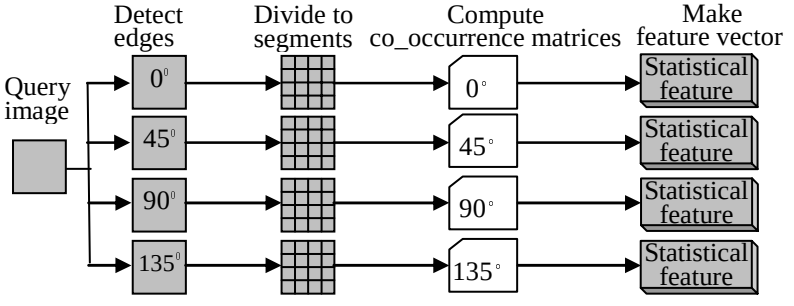


Fig. 2. Composition of edge information and GLCM properties

3 Retrieving Images and Performance Evaluation

The proposed retrieval system consists of four major parts. The first one is that obtains four edge images by using four directions edge operators of Prewitt edge detector. Usually, in order to get clearer edges, a threshold is used in the edge image, but it would lose a part of texture information by the threshold set, so that the proposed system doesn't apply the threshold. Second, the image should be divided into 4, 9 or 16 segments before compute GLCM, in the proposed method we divided every edge image into 16 segments. Third, construct same direction co-occurrence matrix for all segments, and compute statistical features: contrast, entropy, energy and homogeneity. The end, regard the set of statistical feature as feature vector and obtain distance measures using Euclidean distance.

Using GLCM to extract image texture feature is the most typical method in CBIR system. In 2003, Yao proposed a retrieval system using edge detection and GLCM[5] and he applied Sobel operator as a edge detector. To test performance of our system, we compared with Yao's system on same database. Figure 3 shows comparison of retrieval results. In every result, we select the 20 most similar images in order and the first image marked rectangle is the query image. The first row is results of Yao's method; the second row is results of the proposed method. It clearly indicates that this proposed method is better than Yao's method when the query image is same.

Figure 4 shows comparison of precision, it is average precision of 10 classes' image; each class contains 30 similar images. The line marked diamond is Yao's

method, the line marked rectangle is the proposed method. The average precision of Yao's method is 88.2%, the proposed method is 97.9%.

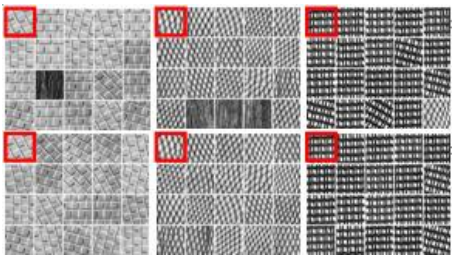


Fig. 3. Comparison of retrieval results

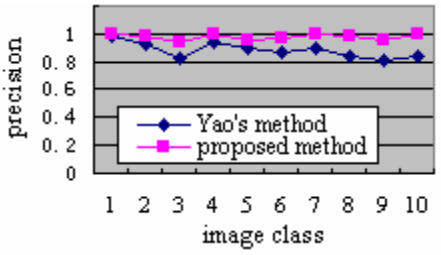


Fig. 4. Comparison of precision

4 Conclusion

In this paper, a new texture-based retrieval method was proposed. Its advantage is that a composition of edge information and gray level co-occurrence matrix properties oriented to four directions was used, adequately to use texture information held in edge image. The results show that the proposed method has high retrieval precision. It gives encouraging results when comparing its retrieval performance to Yao's approach.

References

1. Zhao, W.S., Liu, G.Z., Zhou, Y.T.: Texture image retrieval and similarity matching. Proceedings of 2004 International Conference on Machine Learning and Cybernetics(2004) 4081 - 4084 vol.7
2. Shirahatti, N.V., Barnard, K.: Evaluating image retrieval. Computer Society Conference on Computer Vision and Pattern Recognition(2005)955 - 961 vol. 1
3. Partio, M., Cramariuc, B., Gabbouj, M.: Texture retrieval using ordinal co-occurrence features. Proceedings of the 6th Nordic Signal Processing Symposium(2004)308 – 311
4. Randy Crane: A Simplified Approach to image processing-Classical and Modern Techniques In C. New Jersey:Hewlett-packard Company (1997) 79 – 95
5. Yao, H.Y., Li, B.C.: An efficient approach for texture-based image retrieval[J]. Neural Networks and Signal Processing (2003)1039 – 1043
6. Partio, M., Cramariuc, B., Gabbouj, M., Visa, A.: Rock texture retrieval using gray level co-occurrence matrix. Nordic Signal Processing Symposium (2002)
7. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural Features for Image Classification[J]. IEEE Trans. on Systems, Man,and Cybernetics (1973)610 – 621.
8. Brodatz, P.:Textures: A Photographic Album for Artists and Designers, New York:Dover Publication (1966)

Real-Time Path Planning Strategies for Real World Application Using Random Access Sequence

Jaehyuk Kwak and Joonhong Lim

Department of Electrical Engineering Computer Science , Hanyang University,
1271 Sa-1 dong, Sangnok-gu, Ansan, Kyeonggi-do
jhkwak@aser.hanyang.ac.kr, jhlim@hanyang.ac.kr

Abstract. Many researches on path planning and obstacle avoidance for the fundamentals of mobile robot have been done. Although many solutions help finding a optimal path, those can be applied to real world only under constrained condition, which means that it is difficult to find a universal algorithm. Moreover, a complicated computation to obtain an optimal path induces the time delay so that a robot can not avoid moving obstacles. In this paper, we propose the algorithm of path planning and obstacle avoidance using Random Access Sequence(RAS) methodology. In the proposed scheme, the cell decomposition is make first and cell neighbors are assigned as sequence code, then the path with minimum length is selected.

1 Introduction

Path planning and avoid obstacles are central to navigation strategy of mobile robots. Many researchers have investigated various navigation strategies where a mobile robot has to move from its current position to goal position in known and unknown terrain, respectively. However each proposed strategy is only applied to a specific condition of environment terrain, robot dimension, and shape of obstacles, etc. In the incompletely known domain, robot path planning to find a goal has focused on the completeness rather than the optimality[1]. Usually, a fast replanning is essential in unknown environment[3-4]. Skeleton, cell decomposition, and potential field approaches have been proposed for path planning[5-6]. A recent survey of relevant literature can be found in [2]. Although a lot of algorithms for path finding have been proposed, the most popular solution is A* since it combines the efficiency of heuristic and incremental searches, yet still finds shortest paths algorithm based graph-theory is usually. Moreover Dynamic A*(D*) achieves a large speedup over repeated A*[3] searches by modifying previous search results locally.

Generally, the path finding process is composed of cell decomposition and path selection. In the first process, all space is divided into a small region using their algorithms and then the path to move is selected by A* or D*. From this episode, we can get the shortest and complete path. This means that it is difficult to get both of optimality and completeness by only one method. In this paper, we propose a path planning algorithm for real world mobile robot with various size and shape.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 801 – 804, 2006.

2 Path Planning by Random Access Sequence(RAS)

In the proposed RAS, we use a square shape and set region to cell and move from the center of start square to the center of the next until the target is reached. Initially, the RAS algorithm separates whole regions from each cell of traversable or untraversable. Fig. 1. shows a RAS basic scheme in cell decomposition domain and each cell has eight neighbor cells that neither can reach or not. Similar to A* algorithm, the RAS has a incremental process of numbering but not heuristic. Only using the information of a region not occupied by obstacle, the IDs are added by one successively. The relationship between cells from origin point 0 can be expressed as a hierarchical structure and each circle will be utilized to a ***step*** during numbering process.

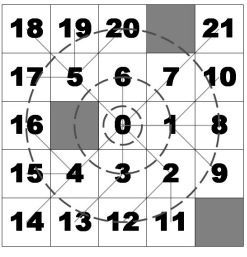


Fig. 1. Basic scheme of RAS

2.1 Notation

To explain the RAS we use the following notations. C denotes the finite set of regions from initial point to destination. k_c is the identified number(ID) of $c \in C$. P_c denotes the center location of region and N_c is the numbers of nearby regions of ID. Since the shape of region is square, the value of N_c is up to 8. $Neigh(c)$ denotes the set of ID of each $c \in C$ in whole space. $Start(k_c)$ is the start ID and defines $Start(k_c) = Neigh(k_{c'})_{biggest}$ $for\ k_{c'} = k_c - 1$. Finally, the density value can be used for various task and expressed to d_c.

2.2 Numbering Process

A numbering process to determine ID of regions is performed by a operation that it starts from a origin point and expand to outside like a swirl. The first step in the origin point is the process of searching neighbored region and give a ID to the traverse region. The next step starts from a lowest number of the former step and performs the same process of find obstacles. Then, it gives a subsequent number from the last sequence number to the region. When all regions have IDs except obstacle,

each cell is connected into hierarchical structure. The IDs from start point can be expressed by

$$\begin{array}{cc} \text{q}^{\text{th}}\ \textbf{\textit{step}} & \\ k_c = Start(k_c) & k_{c'} = Neigh(c'')_{biggest}\ \ for\ k_{c''} = k_c - 1 \\ N_c = n_{k_c} & N_{c'} = n_{k_{c'}} \\ \cdots & \\ Neigh(c) = \{\sum_{0}^{k_c - 1} n_i + 1, ..., \sum_{0}^{k_c} n_i\} & Neigh(c') = \{\sum_{0}^{k_{c'} - 1} n_i + 1, ..., \sum_{0}^{k_{c'}} n_i\} \end{array} \quad (1)$$

2.3 Path Planning Using RAS

A path selection method with RAS is obtained from only ID and neighbor numbers, Assume that $Neigh(c) = \{1,2,3,4,5,6\}, k_c = 0\ (origin)$. That is the start point ID of regions $k_c \in \{1,2,3,4,5,6\}$ is zero. Similarly, if $c \in C$ and $k_c = q$, the start point ID of $Neigh(c)$ is q. From (2), a region number can be written as

$$Neigh(c')_{smallest, k_{c'} = q} = \sum N_c + 1,\ for\ k_c \in \{0,1,...,q-1\} \quad (2)$$

And using (3), the relation for IDs of each cell is explained as

$$\begin{array}{l} Neigh(c')_{smallest} < \sum N_c \le Neigh(c')_{biggest} \\ for\ k_{c'} = q, k_c \in \{0,1,...,q\} \end{array} \quad (3)$$

Only using k_c and N_c, a rout path R_n can be generated by sum operation.

2.4 The Variables of RAS

The most important variables of determining the path are robot size ,region size and density value(d_c). The result path of RAS is only one which has minimum length if a path exists. If a robot size is lager than the empty space of obstacles, a untraversable region may be placed. This problem can be solved when the region has a enough size about robot size. Nevertheless, it is difficult to set the regions exactly and it decrease the probability of finding shortest path. Moreover, smaller region size increases a processing time. Hence we use a density value that is defined as the numbers of neighboring cells which has IDs and not occupied by obstacles and get from

$$d_c = \begin{cases} N_c & if\ step = 0 \\ \sum_{i=0}^{step} \sum_{j=0}^{N_c} N_{c'},\ for\ k_{c'} \in Neigh(c) & else \end{cases} \quad (4)$$

3 Simulation Result

We have conducted path planning experiments using simulated binary terrains. The simulation environment is a binary 1000 x 1000 size world with obstacles by black expressions and it has a obstacle region which has a variant width.

Figure 4 and 5 show the path using A* and RAS, respectively. From this result, both of them have a same path to move to goal from the same initial point and the RAS method showed a more fast execution time compared to A* algorithm.

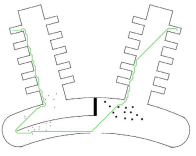


Fig. 4. A path using A*(874.98 msecs)

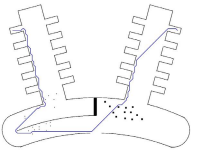


Fig. 5. A path using RAS(2.82 msecs)

4 Conclusion and Future Work

In this paper, we have presented a novel fast replanning method for goal-directed navigation in various environment.. We demonstrated the capability of the RAS by fast planning time in modified environment and the completeness of finding goal position if a path exist. The proposed method guarantees a collision-free path for wall and shows fast planning time. Nevertheless our algorithm cannot guarantee the shortest path, the property of reality, completeness and fast replanning time is important element for real-world application because it can be applied to all conditions. Our further research is that a robot follows the shortest path to goal position using other cell forge and can moves in dynamic environment.

References

1. A. Stentz , " Map-Based Strategies for Robot Navigation in Unknown Environments", Proc. of AAAI , 1996
2. Y.K. Hwang, and N. Ahuja,"Gross motion planning – A survey", ACM Comput. Surv. Vol.24, pp. 219-299, 1992
3. A. Stentz, " The focused D^* algorithm for real-time replanning,", Proc. Of Int. Joint Artificail Intell., pp. 1652-1659, 1995
4. Sven Koenig, Maxim Likhachev, " Fast Replanning for Navigation in Unknown Terrain", IEEE Trans. Robotics, vol. 21, no. 3, pp. 354-363 , 2005
5. S.S. Ge and Y.J. Cui, "New Potential Functions for Mobile Robot Path Planning", IEEE Trans. Robotics, vol. 16, no. 5,pp. 615-620 , 2000
6. Akihiro Kaneshige, Takeshi Akamatsu, Kazuhiro Terashima, "Real Time Path Planning Based on the Potential Method with 3-D Diffusion Process for an Overhead Traveling Crane", Control Conference, vol. 3, pp. 1715-1722, 2004

Multifocus Image Fusion Based on Multiwavelet and Immune Clonal Selection

Xiaohui Yang, Licheng Jiao, Yutao Qi, and Haiyan Jin

Institute of Intelligence Information Processing, Xi'dian University Xi'an 710071
lchjiao@mail.xidian.edu.cn, xhyang_lc@163.com

Abstract. Based on multiwavelet transform and the clonal selection theory in the natural immune system, a novel pixel-level multifocus image fusion optimization algorithm is presented in this paper. Source images are first decomposed into low-frequency coarse information and high-frequency detail information via discrete multiwavelet transform. The high-frequency detail information adopts the absolute-values maximum selection. And then the immune clonal selection is introduced to optimize the weights of fusing the low-frequency four coarse subbands adaptively and separately. Image fusion performances of Daubechies-4 (Db4) scalar wavelet, Geronimo, Hardin and Massopust (GHM) multiwavelets and Chui and Lian (CL) multiwavelets are compared quantitatively, which have the same approximation order. Experimental results show that the proposed image fusion algorithm have clear edges, abundance details and few artificial artifacts.

1 Introduction

Image fusion refers to the techniques that integrate complementary information from multiple sensor data such that the new images are more suitable for the purpose of human visual system (HVS) and the compute-processing tasks. Image fusion has important applications in image analysis and image understanding, such as image segmentation, feature extraction and object recognition.

Pixel-level fusion can hold as many as original dates and then have the highest precision. There have been many image fusion techniques of pixel-level fusion, such as the IHS (Intensity, Hue, Analysis) color model, the PCA (Principal Component Analysis) method, and the wavelet transform (WT) based method [1, 2, 3].

The appropriate transform and fusion rules are two important aspects of image fusion. For the fusion rules in the low-frequency band, there are two appropriate choices: weighted average and estimation methods. The former averages the input low-frequency bands to compose a single low-frequency band. The estimation fusion methods formulate the fusion result in terms of estimated parameters of an imaging model. However, the latter is based on the assumption that the disturbance satisfies a Gaussian distribution, which might mislead the useful signal as disturbance and hence degrade the quality of the fused image. In the high-frequency band, the basic fusion approach is absolute-value maximum selection. It is a fact that the largest absolute values correspond to features in the image such as edges, lines and region boundaries.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 805 – 815, 2006.

Multiwavelets (MWs) are extensions from scalar wavelets (SWs), and it can simultaneously provide perfect reconstruction (PR) while preserving orthogonality, compact support, symmetry, and a higher order of vanishing moments, which cannot be achieved by SWs[4, 5]. As a promising technique, multiwavelet transform (MWT) has been successfully applied to image processing applications with a superior performance as compared to scalar wavelet transform (SWT).

In this paper, a novel pixel-level multifocus image fusion algorithm is presented, which is based on MWT and immune clonal selection (ICS) [6, 7]. By via of MWT, the source images are first decomposed into multiresolution representations with low-frequency coarse information and high-frequency detail information. And then we perform ICS to search the optimal fusion weights for source images adaptively. To multifocus digital camera images, the fused images keep the important information of source images and have more clear boundaries, abundant details and few ringing artifacts. Experimental results compared GHM MW [8, 9] and CL MW [10] with Db4 SW show that the proposed fusion technique is superior to their wavelet counterparts both in visual fidelity and quantitative analysis.

2 Multiwavelets

MWs have more scaling functions $\varphi(t)=[\varphi_0(t),\varphi_1(t)\cdots,\varphi_r(t)]^T$ and mother wavelet functions $\psi(t)=[\psi_0(t),\psi_1(t)\cdots,\psi_r(t)]^T$, where $\phi_l(t)(l=1,2,\cdots,r)$ are orthogonal to each other, and $2^{j/2}\psi_l(t)(2^j t-k)(j,k\in Z,l=1,2,\cdots,r)$ is an orthogonal basis. To implement the MWT, we require a filter bank structure where the low-pass and high-pass filter banks are matrices rather than scalars [11]. The corresponding matrix dilation equations are

$$\begin{aligned}\varphi(t)&=\sqrt{2}\sum\nolimits_n h(n)\varphi(2t-n)\\ \psi(t)&=\sqrt{2}\sum\nolimits_n g(n)\psi(2t-n)\end{aligned}, \tag{1}$$

where $h(n)$ and $g(n)$ are $r\times r$ matrices. A separable MW basis is defined from one-dimensional MW $\psi(t)$ and the corresponding scaling function $\varphi(t)$ which are dilated and translated $\psi_{j,m}(t)=2^{-j/2}\psi(2^{-j}t-k)$ and $\varphi_{j,m}(t)=2^{-j/2}\varphi(2^{-j}t-k)$.

For a given j, if V_j is a close subspace generated by $2^{j/2}\varphi_l(t)(2^j t-k)$ $(j,k\in Z$, $l=1,2,\cdots,r)$ then the subspaces V_j construct an orthogonal multiresolution analysis (MRA). Suppose $f\in V_0$, we have

$$f(t)=\sum_{l=1}^{r}\sum_{k\in Z}c_{l.0,k}\varphi_l(t-k)=\sum_{l=1}^{r}\sum_{k\in Z}c_{l.J_0,k}\varphi_l(2^{J_0/2}t-k). \tag{2}$$

Let $c_{j-1,k} = (c_{1,j,k}, \cdots, c_{N,j,k})^T$, $d_{j-1,k} = (d_{1,j,k}, \cdots, d_{N,j,k})^T$, then the MW decomposition and reconstruction equations can be written as

$$\begin{cases} c_{j-1,k} = \sqrt{2}\sum_n h_n c_{j,2k+n} \\ d_{j-1,k} = \sqrt{2}\sum_n g_n c_{j,2k+n} \\ c_{j,n} = \sqrt{2}\sum_k (h_k c_{j-1,2k+n} + g_k d_{j-1,2k+n}) \end{cases}. \tag{3}$$

As a very important MW system with second approximation, GHM MW was constructed with a fractal interpolation scheme and contains the two scaling functions and the two wavelets. The scaling functions are symmetric and have the short supports [0, 1] and [0, 2] respectively. The wavelet functions are symmetric-antisymmetric. The other useful orthogonal MW system with second approximation is the CL MW, which has both scaling functions and wavelet functions that are symmetric-antisymmetric. The two scaling functions both have the short supports [0, 2] and satisfy Hermite interpolate conditions. Corresponding to each MW system is a matrix-valued multirate filterbank. The lowpass filters h_i and highpass filters g_i of GHM MW and CL MW are in equations (4) and (5) respectively, where $i = 0,1,2,3$.

$$\begin{aligned} h_0 &= \begin{bmatrix} \frac{3\sqrt{2}}{10} & \frac{4}{5} \\ \frac{-1}{20} & \frac{-3\sqrt{2}}{20} \end{bmatrix}, \quad h_1 = \begin{bmatrix} \frac{3\sqrt{2}}{10} & 0 \\ \frac{9}{20} & \frac{\sqrt{2}}{2} \end{bmatrix}, \quad h_2 = \begin{bmatrix} 0 & 0 \\ \frac{9}{20} & \frac{-3\sqrt{2}}{20} \end{bmatrix}, \quad h_3 = \begin{bmatrix} 0 & 0 \\ \frac{-1}{20} & 0 \end{bmatrix} \\ g_0 &= \begin{bmatrix} \frac{-1}{20} & \frac{-3\sqrt{2}}{20} \\ \frac{\sqrt{2}}{20} & \frac{3}{10} \end{bmatrix}, \quad g_1 = \begin{bmatrix} \frac{9}{20} & \frac{-\sqrt{2}}{2} \\ \frac{-9\sqrt{2}}{20} & 0 \end{bmatrix}, \quad g_2 = \begin{bmatrix} \frac{9}{20} & \frac{-3\sqrt{2}}{20} \\ \frac{9\sqrt{2}}{20} & \frac{-3}{10} \end{bmatrix}, \quad g_3 = \begin{bmatrix} \frac{-1}{20} & 0 \\ \frac{\sqrt{2}}{20} & 0 \end{bmatrix} \end{aligned} \tag{4}$$

$$\begin{aligned} h_0 &= \begin{bmatrix} \frac{1}{2\sqrt{2}} & \frac{-1}{2\sqrt{2}} \\ \frac{\sqrt{7}}{4\sqrt{2}} & \frac{-\sqrt{7}}{4\sqrt{2}} \end{bmatrix}, \quad h_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{2\sqrt{2}} \end{bmatrix}, \quad h_2 = \begin{bmatrix} \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ \frac{\sqrt{7}}{4\sqrt{2}} & \frac{-\sqrt{7}}{4\sqrt{2}} \end{bmatrix} \\ g_0 &= \begin{bmatrix} \frac{1}{2\sqrt{2}} & \frac{-1}{2\sqrt{2}} \\ \frac{-\sqrt{7}}{4\sqrt{2}} & \frac{\sqrt{7}}{4\sqrt{2}} \end{bmatrix}, \quad g_1 = \begin{bmatrix} \frac{-1}{\sqrt{2}} & 0 \\ 0 & \frac{\sqrt{7}}{2\sqrt{2}} \end{bmatrix}, \quad g_2 = \begin{bmatrix} \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ \frac{\sqrt{7}}{4\sqrt{2}} & \frac{\sqrt{7}}{4\sqrt{2}} \end{bmatrix}. \end{aligned} \tag{5}$$

3 Image Fusion Based on MW and ICS

The four blocks of the lowpass subbands of MW domain have common been combined using the same fusion rule, and the fusion weights are adopted by experiments. It is clear that this approach does not take into account the spectral dissimilarity of these blocks. To different source images, the optimal weights actually should not be the same. In this section, we give an optimization algorithm of image fusion based on the immune clonal selection (ICS). This novel technique treats the four subbands separately and via ICS to obtain the optimal weights adaptively.

3.1 Preprocessing of Multiwavelet

The lowpass filters and highpass filters of MW system are $n \times n$ matrix, and during the convolution step they must multiply vector instead of scalars. This means that multifilter banks need n input rows. The most obvious way to get two input rows from a scalar signal is to repeat the signal. It introduces oversampling of the date by a factor of two. To image signal, the oversampling factor increases to four and introduce too much redundancy and complex computation [12, 13].

In this section, we adopt the critically sampled scheme to prefilter the given scalar signal. This preprocessing algorithm also maintains a critically sampled representation: if the data enters at rate R , reprocessing yields two streams at rate $R/2$ for input to the multifilter, which produces four output streams, each at a rate $R/4$. In practice all signals have finite length, so we adopt symmetric extension technique for filtering such signals at their boundaries to preserve critical sampling and continuity. Two-level MW decomposition procedure is shown in figure 1:

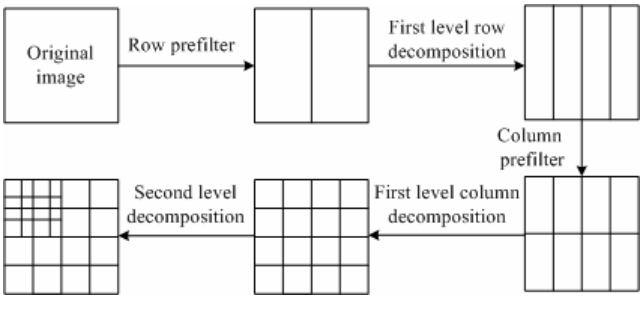


Fig. 1. Multiwavelet decomposition based on prefilter

The corresponding prefilter P_{re} and postfilter P_{ost} of GHM multiwavelet and CL multiwavelet are in equations (6) and (7) respectively:

$$P_{re}(0)=\begin{bmatrix}\frac{3}{8\sqrt{2}} & \frac{10}{8\sqrt{2}}\\ 0 & 0\end{bmatrix},\quad P_{re}(-1)=\begin{bmatrix}\frac{3}{8\sqrt{2}} & 0\\ 1 & 0\end{bmatrix}$$
$$P_{ost}(1)=\begin{bmatrix}0 & 1\\ 0 & -\frac{3}{10}\end{bmatrix},\quad P_{ost}(0)=\begin{bmatrix}0 & 0\\ \frac{4\sqrt{2}}{5} & -\frac{3}{10}\end{bmatrix}, \tag{6}$$

$$P_{re}(0)=\begin{bmatrix}\frac{1}{4} & \frac{1}{4}\\ \frac{1}{1+\sqrt{7}} & \frac{-1}{1+\sqrt{7}}\end{bmatrix},\quad P_{ost}(0)=\begin{bmatrix}2 & \frac{1+\sqrt{7}}{2}\\ 2 & \frac{-(1+\sqrt{7})}{2}\end{bmatrix}. \tag{7}$$

Figure 2 shows a single-level decomposition of Clock image using discrete GHM MW (left) and CL MW (middle). Unlike scalar wavelets, each decomposition level consists of 16 subbands (right). Especially, the lowpass subband consists of four sub-blocks characterized by different spectral properties, make L -level decomposition using discrete MW look like $L+1$ -level decomposition using DSW. Therefore, we should not only adopt different fusion rules to lowpass subband and highpass subband, but also treat the four sub-blocks of lowpass subband separately, and optimization the fusion weights adaptively.

Fig. 2. Single-level decomposition of Clock image

3.2 Immune Clonal Selection Optimization

As a novel artificial intelligent optimization technique, the artificial immune system (AIS) aim at using ideas gleaned from immunology in order to develop systems capable of performing different tasks in various areas of research. The clonal selection functioning of the immune system can be interpreted as a remarkable microcosm of Charles Darwin's law of evolution, with the three major principles of diversity, variation and natural selection.

The clonal selection algorithm is used by the natural immune system to define the basic features of an immune response to an antigenic stimulus. The main features of the clonal selection theory are: generation of new random genetic changes subsequently expressed as diverse antibody patterns by a form of accelerated somatic mutation; phenotypic restriction and retention of one pattern to one differentiated cell

(clone); proliferation and differentiation on contact of cells with antigens. It establishes the idea that only those cells that recognize the antigens are selected to proliferate. The selected cells are subject to an affinity maturation process, which improves their affinity to the selective antigens. Random changes are introduced and will lead to an increase in the affinity of the antibody. It is these high-affinity variants which are then selected to enter the pool of memory cells. Those cells with low affinity receptors must be efficiently eliminated become anergic or be edited, so that they do not significantly contribute to the pool of memory cells.

3.3 MW+ICS Fusion Algorithm

In this section, we propose the fusion algorithm based on MW and the ICS named MW+ICS optimization fusion algorithm.

We consider the characteristics of coefficients on each multiresolution decomposition level and each subband, and adopt different fusion rules to lowpass subband and detail subbands. More significance, we treat the four coarse subbands separately and via optimization of CSA to obtain the fusion weights adaptively. Figure 5 illustrates a single analysis/synthesis stage of MW processing.

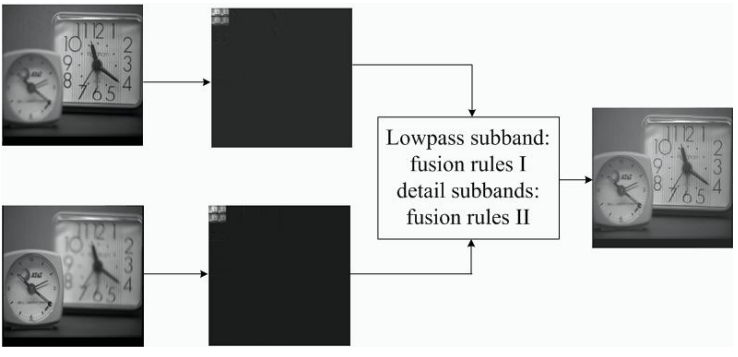


Fig. 5. Image fusion processing using GHM multiwavelet decomposition

The cost function defined as the values of the distortion rate and as the affinity function of ICS: $A = \| f - f_R \|_2^2$, where $\| f - f_R \|_2^2$ is the mean square error of desire image and the fused image. During the process of searching optimization weights, we introduce the elitist preserved definition to keep the weights corresponding to the current best affinity function and save the memory space.

Definition (Elitist preserved). Suppose that $S^* = \{S^* : f(S_l^*) = \min(f(S_l^*)), l = \lg 2(N), \cdots, 2, 1\}$, where S_l^*, S_{l-1}^* are the sets (memory population) of optimal directions on l level and $l+1$ level, $f(S_l^*)$ and $f(S_{l-1}^*)$ are the corresponding values of object function. If $f(S_l^*) > f(S_{l-1}^*)$, then $S_l^* := S_{l-1}^*$ and $f(S_l^*) := f(S_{l-1}^*)$.

Without of generalization, we only focus on two source image. There is the same to many source images. Suppose all the source images are registrated, that is, each image being aligned to the same pixel position. If the sensors to be fused are not perfectly aligned, any corrections to the alignment can be made through a low-latency image warp function. The fusion algorithm is as follows:

Step1. Loading source images;
Step2. Performing critical sampling prefiltering;
Step3. Performing three level multiwavelet transform. Suppose that the source images are I_1 and I_2, we have the approximation sub-band on the third level and the subimage series of detail sub-band on each level. And they denote $\left\{(A_{I_i}, D_{I_i}^{k,m}) \mid i = 1,2;\ 1 \le k \le 3; 1 \le m \le 4\right\}$ respectively;

Step4. Performing the fusion rule of absolute-values maximum to combine corresponding detail sub-band images $D_{I_i}^{k,m}$:

$$\{F_{cH}, F_{cV}, F_{cD}\} = \begin{cases} \{cH_1, cV_1, cD_1\}, & if \ \ |cH_1| + |cV_1| + |cD_1| > |cH_2| + |cV_2| + |cD_2| \\ \{cH_2, cV_2, cD_2\}, & otherwise \end{cases} \tag{8}$$

where $\{cH_1, cV_1, cD_1\}$ and $\{cH_2, cV_2, cD_2\}$ are the corresponding wavelet coefficient matrixes of the three high-frequency sub-band of source images I_1 and I_2;

Step5. Performing the ICS to search the optimal fusion weights to the corresponding approximation sub-images on the last decomposition level adaptively,

$$F_{cA} = \alpha cI_1 + (1-\alpha) cI_2, \tag{9}$$

where cI_1, cI_2 and F_{cA} denote the approximation subimages of source images I_1, I_2 and the synthesis image, $\alpha (0 \le \alpha \le 1)$ is the resulting weights. The corresponding ICS sub-algorithm as following and shown in figure 6:

Step5.1. Initialization. Pre-select the weights in (0, 1) and denoting them initial group individuals with the size of $i = 9$. Let the clone generations is ten.
Step5.2. Calculating the affinity values of the initial group individuals and storing them in memory cell in sort;
Step5.3. Clone. Cloning the initial group. Suppose the clone size is three, that is, each individual is cloned to three child individuals by random changes around the father individual. And then we calculating the affinity values of the child individuals;
Step5.4. Aberrance. Compare the three child individuals with the corresponding father individual, based on an affinity measure. If the affinity value of the child is larger than its father counterpart, then the former replaces the latter and be kept to the memory cell $S := S_{opt}$. Otherwise, the father individual is kept;
Step5.5. Reselection. Return to step5.3 and repeat the optimization procedure until satisfy the stop condition.

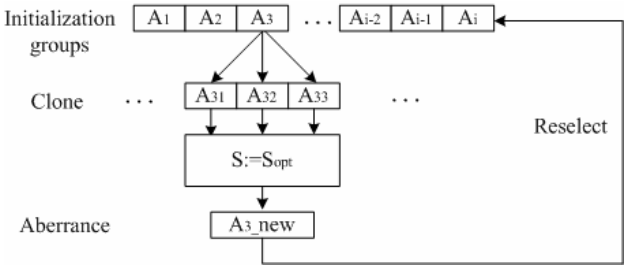


Fig. 6. Block diagram of the clonal selection algorithm

Step6. Substitute the resulting optimal weights to equation (12) and fusion the appreciation subimages:
Step7. Performing inverse multiwavelet transform to the fused sub-image serials;
Step8. Performing postfiltering on the synthesis image and obtain the fused image.

4 Experiments

In this section, we adopt multifocus source images to test our fusion effects. All the source images are registrated gray images (0~255), with the size of 256×256 and 512×512. The decomposition levels of GHM MW and CL MW are three, and the Db4 is four.

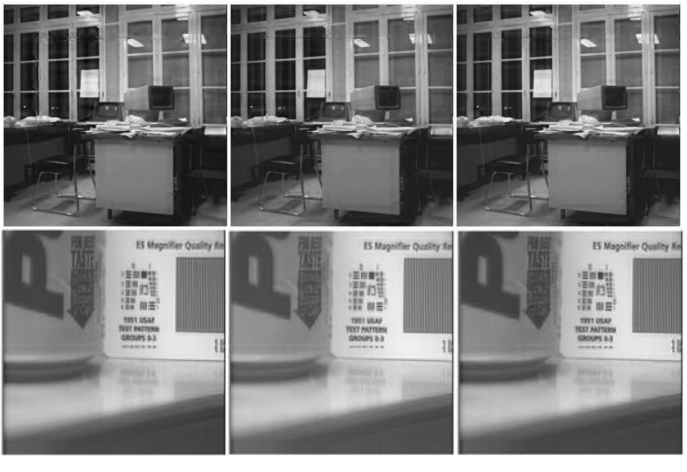


Fig. 7. Comparisons of Pepsi image

Figure 7 gives the fusion results of the office image and pepsi image based on db4+ICS (left), GHM+ICS (middile) and CL+ICS (right) respectively. From figure 7 we can conclusion that the fusion images based on MW+ICS are superior to SW+ICS.

There are few invisible ringing artifacts and more clearly. And the fusion results based on CL+ICS method is more efficient than based on GHM+ICS method.

The subjective visual apperceive is only one evaluation rules and can be effected in a certain extent by vision psychology factor. The following gives the quantitative rules.

1. Root Mean Square Error

The root mean square error (RMSE) between the synthesis image and the desired image is defined as:

$$RMSE = \sqrt{\frac{1}{NM}\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}[I_I(i,j) - I_F(i,j)]^2}\ , \tag{10}$$

where M and N are the sizes of images, $I_I(i,j)$ and $I_F(i,j)$ are the pixel of source image and fused image respectively. The smaller the RMSE is, the better the fusion effect is.

2. Standard deviation

Standard deviation (STD) is an important index of image traffic and reflect the dispersedly degree of gray value and gray mean value. The larger the STD is, the more dispersive the gray levels distribute. And the appearance probability of all the gray levels goes equal and the containing traffic goes the largest.

$$\sigma = \sqrt{\frac{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}[I_I(i,j) - I_F(i,j)]^2}{NM}}\ , \tag{11}$$

where $I_I(i,j)$ and $I_F(i,j)$ denote the gray value of desired image and fused image on (i,j) respectively. The smaller σ is, the better the fusion is.

3. Information entropy

Information entropy is an important weigh index of richness degree of image information. It shows the number of the average traffic contains in image. The larger the entropy is, the larger the traffic is. Based on Shannon information theory, the entropy is defined as:

$$E = -\sum_{i=0}^{255} P_i \log_2 P_i\,, \tag{12}$$

where P_i is the probability of image pixel gray values equal to i.

From table 1 and table 2, we have the conclusion that the fusion results based on Db4 inferior to their MW counterparts. The fusion based on our proposed MW+ICS are better than the wavelet contourparts. The best of all is the CL+ICS fusion, which has the larger information entropy and small reconstruction error.

Table 1. Comparison of fusion performance on Office image

Office image		STD	RMSE	ENTROPY
Original A		59.5480	25.9810	5.0052
Original B		63.4370	19.5280	5.0931
Desire image		48.9760	——	5.2865
Fused image	Db4	53.2500	11.1355	5.1705
	CL	50.1890	8.3916	5.1796
	GHM	51.8700	10.6380	5.1794
	Db4+ICS	51.1995	10.8187	5.1744
	CL+ICS	49.3097	8.3089	5.1856
	GHM+ICS	50.1007	10.344	5.1824

Table 2. Comparison of fusion performance on Pepsi image

Pepsi image		STD	RMSE	ENTROPY
Original A		43.8158	11.2750	4.9130
Original B		44.9985	4.8684	4.9272
Desire image		39.5315	——	4.9388
Fused image	Db4	42.9255	3.6324	4.9228
	CL	41.1890	8.3916	4.9267
	GHM	41.9510	3.7163	4.9254
	Db4+ICS	41.0290	3.5109	4.9271
	CL+ICS	40.4580	3.3331	4.9357
	GHM+ICS	40.9673	3.6203	4.9310

5 Conclusions

This paper presents a novel fusion algorithm, which is based on multiwavelet transform and immune clonal selection optimization. As a novel intelligence optimization technique, the immune clonal selection technique is introduced into image fusion to obtain the optimal fusion weights adaptively. Experimental results show that the proposed approach has improvements in visual fidelity and quantitative analysis. How to solve the fusion problem of remote images without desired compared images is our future work.

References

1. Nunez, J., Otazu, X., Fors, O., Prades, A., Pala, V., Arbiol, R.: Multiresolution based image fusion with additive wavelets decomposition. IEEE Trans. Geosci. Remote Sensing, 37 (1999) 1204-1211
2. Zhang, Z., Blum, R. S.: A categorization of multiscale decomposition-based image fusion schemes with a performance study for a digital camera application. Proceedings of the IEEE. 87 (1999) 1315–1326

3. Kazemi, K., Moghaddam, H. A.: Fusion of multifocus images using discrete multiwavelet transform. IEEE Conf. on Multisensor Fusion and Int. for Int. Systems. (2003) 167-172
4. Salesnick, I.: Multiwavelet bases with extra approximation properties. IEEE Trans. on Signal Processing, 46 (1998) 2898-2908
5. Strela, V.: Multiwavelets: Theory and Applications. Ph.D Thesis, MIT. (1996)
6. De Castro, L. N., Von Zuben, F. J.: The Clonal Selection Algorithm with Engineering Applications. Proc. of GECCO'00, Workshop on Artificial Immune Systems and Their Applications. (2000) 36-37
7. Liu, R. C., Du, H. F., Jiao, L. C., Immunity clonal strategies. Proc of ICCIMA'03. The Fifth International Conference on Computational Intelligence and Multimedia Applications.
8. Geromino, J. S., Hardin, D. P., Massopust, P. R.: Fractal functions and wavelet expansions based on several scaling functions. J. Approx. Theory. 78 (1994) 373–401
9. Donovan, G., Geromino, J. S., Hardin, D. P., Massopust, P. R.: Construction of orthogonal wavelrts using fraction interpolation functions. SIAM J. Math. Anal. 27 (1996) 1158–1192
10. Chui, C. K., Lian, J. A.: A study of orthonormal multiwavelets. Applied Numerical Mathematics. 20 (1996) 273–298
11. Mallat, S.: A theory for multiresolution signal decomposition: wavelet representation. IEEE Trans. Pattern Anal. Machine Intell. 11 (1989) 674–693
12. Strela, V., Heller, P. N., Strang, G., Topiwala, P., Heil, C.: The application of multiwavelet filterbanks to image processing. IEEE Trans. on Image Processing, 8 (1999) 548-563
13. Attakitmongcol, K., Hardin, D. P., Wilkes, D. M.: Multiwavelet Prefilters-Part II: Optimal Orthogonal Prefilters. IEEE Trans. on Image Processing, 10 (2001) 1476-1487

Numerical Study on Propagation of Explosion Wave in H_2-O_2 Mixtures

Cheng Wang, Jianguo Ning, and Juan Lei

State Key Laboratory of Explosion Science and Technology, Beijing Institute of Technology, Beijing 100081, P.R.China
wangcheng@bit.edu.cn, jgning@bit.edu.cn

Abstract. Two-dimensional computations of the propagation of the explosion wave in a mixture of hydrogen and oxygen are performed using WENO finite difference scheme with fifth order accuracy and two-step chemical reaction model. Numerical results show that obstacles setting repeatedly along the path of the flame propagation process obviously can accelerate the flame. The mechanism of the accelerating effect is attributed to the positive feedback of the turbulence region induced by obstacles for combustion process.

1 Introduction

Flame acceleration and deflagration-to-detonation transition (DDT) are important phenomena in severe accidents because they can largely influence the maximum loads from hydrogen combustion sequences and the consequential structural damage. In early 1980's, much more attention is drawn to study the mechanisms of obstacles accelerating flame propagation and of pressure wave intensification. Lee et al [1] analyzed the phenomena of flame acceleration in an enclosed duct with H_2 and C_2H_2. Their experimental results indicate that the turbulent effect induced by obstacles can significantly accelerate flame. In this paper, a series of numerical experiments on hydrogen and oxygen mixture are conducted to study flame propagation in ducts with or without obstacles adopting WENO (Weighted essentially non-oscillatory) scheme with fifth order accuracy and two-step chemical reaction mode, and the propagation characteristics of flame are obtained.

2 Numerical Method

Numerical studies have been performed by two-dimensional Euler equations with two-step reaction model.

$$\frac{\partial U}{\partial t}+\frac{\partial F}{\partial x}+\frac{\partial G}{\partial y}+\phi=0 \ . \tag{1}$$

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 816 – 819, 2006.

$$U=\begin{pmatrix}\rho\\ \rho u\\ \rho v\\ \rho e\\ \rho\alpha\\ \rho\beta\end{pmatrix}\quad F=\begin{pmatrix}\rho u\\ \rho u^2+p\\ \rho uv\\ (\rho e+p)u\\ \rho\alpha u\\ \rho\beta u\end{pmatrix}\quad G=\begin{pmatrix}\rho v\\ \rho uv\\ \rho v^2+p\\ (\rho e+p)v\\ \rho\alpha v\\ \rho\beta v\end{pmatrix}\quad \phi=\begin{pmatrix}0\\ 0\\ 0\\ 0\\ \rho\omega_\alpha\\ \rho\omega_\beta\end{pmatrix}. \tag{2}$$

$$p=\rho RT\ . \tag{3}$$

$$e=\frac{RT}{\gamma-1}+\beta q+\frac{1}{2}(u^2+v^2)\ . \tag{4}$$

$$\omega_\alpha=-K_1 p\exp(-E_1/RT)\ . \tag{5}$$

$$\omega_\beta=\begin{cases}-K_2 p^2[\beta^2\exp(\dfrac{-E_2}{RT})-(1-\beta)^2\exp(-\dfrac{E_2+q}{RT})] & (\alpha\le 0)\\ 0 \quad (\alpha>0)\end{cases}. \tag{6}$$

where u and v are velocity in the x and y directions, e is total energy, q is reaction heat per unit mass, ρ is mass density, p is pressure, R is gas constant, T is temperature, and γ is adiabatic index. ω_α and ω_β are the empirical parameters of two-step reaction mode, α and β are reaction progress variables, and K_1, K_2, E_1, E_2 denote the pre-exponential factors and activation energies [3].

The space difference is calculated using fifth-order WENO scheme [2] and time difference with the third-order TVD Runger-Kutta method.

3 Numerical Results

In this paper, the flame propagation of premixed H_2-O_2 mixture is simulated. The duct is 240mm long and with a diameter of 80mm, and is enclosed at both ends. Six cases are simulated in the paper, including ducts with and without obstacles. Three orifice plates are located as obstacles, and the spacing of each other is 50mm. The blockage ratio is 0.96, 0.84, 0.64, 0.5 and 0.3, respectively. Blockage ratio can be defined as.

$$B=(R^2-r^2)/R^2\ . \tag{7}$$

where r and R are inner and outer diameters of plate, respectively.

Flame velocity versus blockage ratio is shown in Fig.1(a), Fig.1(b) illustrates flame velocity versus propagation distance in duct with and without obstacles, and Fig.2 demonstrates the flame propagation process of H_2-O_2 mixture in ducts, and N is calculation step.

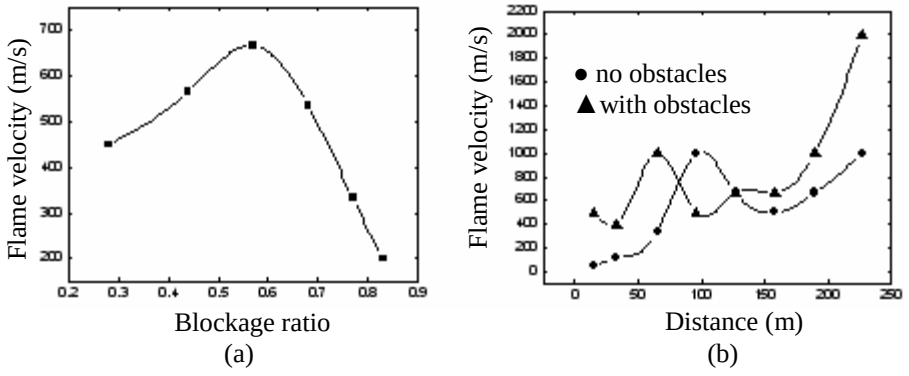


Fig. 1. (a) Flame speed versus blockage ratio, (b) Flame velocity versus propagation distance

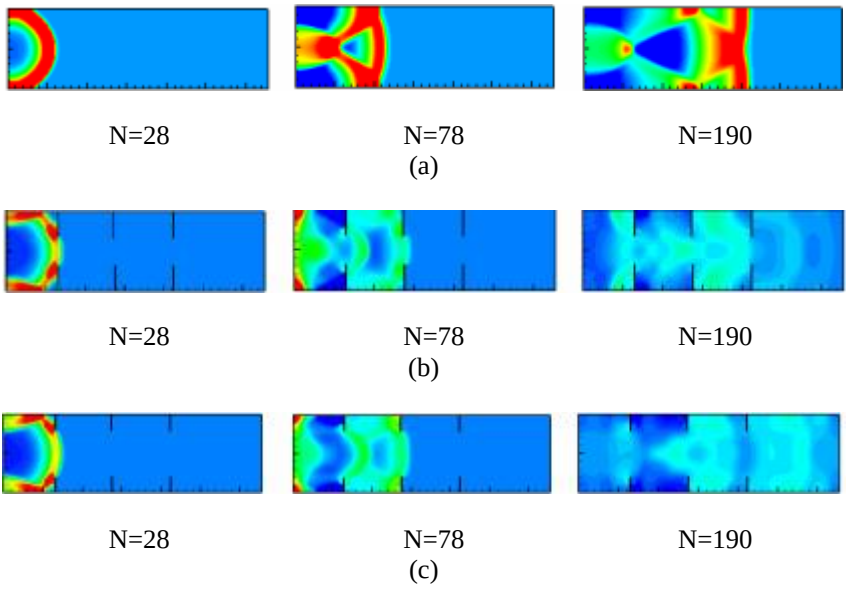


Fig. 2. Flame propagation process of H_2-O_2 mixture in ducts under different conditions of (a) no obstacles, (b) obstacles with a blockage ratio of 0.64, (c) obstacles with a blockage ratio of 0.84

It can be concluded from Fig.1 and Fig.2 that flame velocities in ducts with and without obstacles increase linearly at the early stage of flame propagation process. But it increases smoothly in duct without obstacles, and oscillate severely in ducts with obstacles. The reason is that the existence of obstacles turns laminar flame into turbulent flame and increases the combustion area, which leads to a sudden acceleration of flame. Flame velocity generally increase linearly with blockage ratio, however, as blockage ration gets to the certain value, flame velocity begins to decrease or even flame quenches sometimes. The reason is that higher blockage ratio can reduce

the flow velocity of gas behind obstacles, and hence make flame velocity decrease. Although the flame velocity still increases as blockage ratio rises, the increment is lower than the decrement of gas flow velocity, with a total effect of decreasing flame velocity [1].

Interaction between turbulence and flame is the main factor that leads to flame acceleration, so it is feasible that placing obstacles can help to realize the transition form deflagration to detonation. The results of numerical simulation agree well with previous research and have good repeatability, which proves that numerical simulation is effective to predict the condition of deflagration to detonation in some cases.

4 Conclusion

The flame propagation process of premixed H_2-O_2 mixture in an enclosed duct is simulated. Numerical results suggest that the existence of obstacles turns the original laminar flame into turbulent flame and increases the combustion area, which leads to a sudden acceleration of flame. Flame acceleration is due to integrated effects of unburned gas in front of flame front heated by leading compression wave and of positive feedback between obstacle-induced turbulent zone and the combustion.

Acknowledgments

The work is supported by the National Natural Science Foundation of China (10572026).

References

1. Lee, J. H. S., Knystautas, R., Chan, C. K.: Turbulent Flame Propagation in Obstacle-Filled Tubes. In 20th Symposium (International) on Combustion, The Combustion Institute, Pittsburg (1984) 1663-1672
2. Chi-Wang Shu: Essentially Non-oscillatory and Weighted Essentially Non-oscillatory Schemes for Hyperbolic conversation Laws. NASA/CR-97-206253, ICASE Report No. 97-65 (1997) 1-78
3. Korobeinikov, V., Levin, V. A., Markov, V. V., Chernyi, G.G.: Propagation of Blast Wave in A Combusting Gas. Acta Astronautica. 17 (1972) 529-535

Classification of Online Discussions Via Content and Participation

Victor Cheng, Chi-sum Yeung, and Chun-hung Li

Department of Computer Science, Hong Kong Baptist University, Hong Kong
{victor, csyeung, chli}@comp.hkbu.edu.hk

Abstract. Web forums and online communities are becoming increasingly important information sources. While there are significant research works on classification of Web pages, relatively less is known on classification of online discussions. By observing the special nature of user participation in web communities, we propose classification methods based on user participation and text content of online discussions. Support vector machines have been employed in this study to classify and analyze discussions based on content and participation. It is found that the accuracy of using participation as classification features can be very high in certain communities. The use of high-dimensional classifier can be effective in enhancing retrieval and classification of online discussion topics.

Keywords: Online community, support vector machine, information retrieval, feature extraction, classification.

1 Introduction

Online forums are forums that utilize computers and networks as mediators linking people from different areas to communicate and share their idea and/or information with each other. Now they have become an important information repository covering nearly all knowledge domains contributed by thousands of millions Internet users. They span from very general topics to extremely specific items. However, the information is not effectively utilized because current search engines such as Google and Yahoo are not designed [1],[2] to address forums searching.

1.1 Characteristics of Online Forums

To unleash the information stored in online forum, the availability of efficient information retrieval (IR) techniques are significant. However, there are relatively less researches on forum information retrieval compared with that of web pages. Since most messages in forums are posted by ordinary users, they are usually short, incomplete, diverse, informal, and it is less likely that there are inbound hyperlinks to them from other web sites. Interestingly, the cooperative power of these messages is great! They cover general to specific areas of different topics. Many of them even have information that cannot get from ordinary web sites.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 820 – 828, 2006.

Apart from lacking of hyperlinks, the organization of forum messages also makes IR difficult. Many categories of forums are poorly organized and even wrongly organized. Wrongly placed messages are common in most forums. For example, a message about power cable placed in an Audio CD thread may cause this message cannot be found easily. Another issue is the nature of messages. Some messages should be placed in multiple categories and at the same time there is situation that it is difficult to find a category to place messages. To resolve this problem, many forums are equipped with the "General", "Others", or "Misc" category. This makes the problem more complex because most users tend to place their messages there, for the sake of convenience or other reasons, and this category eventually hold a lot of different messages mixing together without proper classification.

In this paper we propose forum threads classification based on user participation and text content. Although the messages in a forum are not hyperlinked, additional attributes, poster names, provide implicit linkage information between messages. The idea is "important poster creates important messages for threads", particularly in very specific domains. This relationship is similar to the functions of hyperlinks for web pages, "popular and important web pages are usually highly hyperlinked by other important pages". Support vector machine (SVM) [4],[5] have been employed in this study to classify and analyze online forums based on content and participation. It is found that the accuracy of using poster participation as attributes can give classification very high accuracy in certain communities.

The contents of this paper are organized as follow. Section 2 describes the similarities of threads measured by text and posters of forums. The SVM techniques for classification are covered in Section 3. Section 4 discusses the application of the techniques for classifying threads of a popular web forum R concerning audio video equipment. Finally, a conclusion is given in Section 5.

2 Threads Similarities of Forums

Before discussing the threads classification of online forums, the following entities are defined.

- **Forum:** A cyber space utilizes computers and networks as mediators linking people from different areas to communicate and share their idea and/or information with each other.
- **Category:** A collection of threads which share the same theme.
- **Post /Message:** A message created by a poster.
- **Thread:** A collect of posts which share essentially the same topic.
- **Poster:** Author of a post message. Each poster may initialize or replay a number of threads of their interests. They may participate in different threads or different forums freely.
- **Subject:** A collection of word which is the topic of a thread and it can be regarded as a general summary of the thread.

In our study, posters and text of messages are separately used as attributes in similarity measure between threads. It is found that each of them have their own merits in certain situations.

2.1 Similarity Measure for Threads Using Text

The similarity measure of threads using text follows the techniques used in document classification; a thread is analogous with a document. With the Vector Space Model (VSM) [6], a thread is represented by a terms vector with non-zero entities corresponding to the presence of the terms in messages of the thread. A forum is then can be represented by a threads-by-terms matrix.

Similar to document classification, *TFIDF* [7] is employed in our study. It sticks out key terms of threads and gives higher weight to them. The *TFIDF* of a term in a thread vector is given by

$$TFIDF_i = TF_i \times IDF_i . \tag{1}$$

Where TF_i is the frequency of occurrence of term *i* in the thread,
and IDF_i = 1/log(Number of threads containing the term *i*)

For a forum with *l* terms, a thread vector, T_m, is denoted by

$$(TFIDF_{m,1}, TFIDF_{m,2}, \ldots\ldots, TFIDF_{m,l}).$$

The similarity, sim, between two threads, T_m and T_n, is given by

$$sim = \frac{\left(\sum_{i=1}^{l} TFIDF_{m,i} \times TFIDF_{n,i} \right)^{1/2}}{\left| T_m \right\| T_n \right|} . \tag{2}$$

2.2 Similarity Measure for Threads Using Posters

Although forums lack of hyperlinks, the posters of an online forum give an implicit information of relationship between threads. It is because a thread in a forum usually condense a group of posters sharing similar interest. Imagine a thread discussing audio cables in a Hi-Fi forum. The participants are believed to have interest in audio cables. If these group of people appear in another thread, this thread is believed to have the subject about audio cables. This phenomenon is due to the fact that most people are interested in few areas, especially for very specific areas. It is not easy to find a people have deep knowledge in all domains. As a result, the appearance of messages from a group of people in a thread has implication on the topic discussed. When the same group of people discuss together in two threads, such two thread are likely to be similar as both thread exhibit certain features that attract the same group of people who only nterested in narrow range of topic. Hence, the presence of common posters of two threads gives a measure on the similarity between these threads. The threads similariy has similar formulation as that given in Section 2.1, with each distinct poster is analogous to a distinct term. TFIDF is also used to identify important posters and grant them higher weights. The TFIDF for posters *i* is interpreted as follow.

$$TFIDF_i = TF_i \times IDF_i . \tag{3}$$

Where TF_i is the frequency of occurrence of poster *i* in the thread,
And IDF_i = 1/log(Number of threads having the poster *i*)

For a forum with p posters, a thread vector, T_m, is denoted by

$$(TFIDF_{m,1}, TFIDF_{m,2}, \ldots\ldots, TFIDF_{m,p}).$$

The similarity, sim, between two threads, T_m and T_n, using posters as attributes is given by

$$sim = \frac{\left(\sum_{i=1}^{p} TFIDF_{m,i} \times TFIDF_{n,i}\right)^{1/2}}{\left|T_m\right\|T_n\right|}. \tag{4}$$

Threads with a few posters should be removed because similarity between threads due to co-occurrence of posters is not operative. In our study, it is also fuond that posters of threads with general topics are less informative in measuring threads similarity. It is because they attract people of different knowledge domains and hence the participants of these threads do not show any bias to any knowledge domains and thus they usually have high similarity to other threads. For example, the thread discussing audio speakers will attract posters of all Hi-Fi aspects because audio speakers are common equipment and almost every Hi-Fi player have interest in it. As a result, this thread has high similarity to other threads.

3 SVM Classification

Classification of threads is done with Support Vector Machine (SVM) developed by V. Vapnik [8]. SVM has a solid foundation based on Statistical Learning Theory [9] and it usually gives good generalization performance. The idea of SVM classification is similar to that of perception [10]. It finds a linear separation boundary/hyperplane described by $\boldsymbol{w}^T \boldsymbol{x} + \boldsymbol{b} = 0$ to classify samples into two classes. As there are infinite hyperplanes separating samples into two classes (assume samples are linear separable), SVM searches one that giving maximal margin, the distance to the closest training sample is maximal. This property enables the techniques to find a separating plane with good generalization.

Let $\boldsymbol{X}=\{(\boldsymbol{x_i}, c_i)$, i=1,….,n, $\boldsymbol{x_i} \in R^m$, $c_i \in \{-1, +1\}\}$ denotes a set of training samples. Assume $\boldsymbol{X}$ is linearly separable, there exist a hyperplane $\boldsymbol{w}^T \boldsymbol{x} + \boldsymbol{b} = 0$ such that sign$(\boldsymbol{w}^T \boldsymbol{x_i} + \boldsymbol{b})= c_i$ for all i. Let $\boldsymbol{x_i}$ be the sample with minimum distance to the plane. Define the margin under this configuration to be

$$\text{margin} = \frac{\left|w^T x_i + b\right|}{\left\|w\right\|}. \tag{5}$$

By scaling $\boldsymbol{w}$ and $\boldsymbol{b}$ so that the minimum distance from $\boldsymbol{x_i}$ to the plane be 1, a hyperplane with maximum margin can be found by solving the optimization problem

$$\text{Minimize } \frac{1}{2} w^T w \tag{6}$$

with constraint: $c_i(w^T x_i + b) \geq 1, \quad i = 1,2,...,n$.

Using the Lagrange multipliers and Karush-Kuhn-Tucker conditions, the problem can be transformed into a dual form

$$\text{Maxmize } L(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j c_i c_j x_i x_j \tag{7}$$

With constraint: $\alpha_i \geq 0, i = 1,...,n$ and $\sum_{i=1}^{n} \alpha_i c_i = 0$.

Equation (7) is a quadratic optimization problem and it can be solved with quadratic programming routines. Once all α_i are obtained, the optimal hyperplane is determined by

$$w = \sum_{i=1}^{n} \alpha_i c_i x_i \tag{8}$$

$$b = \text{mean of } (\frac{1}{c_k} - w^T x_k) \text{ for } 1 \leq k \leq n \text{ with } \alpha_k \neq 0 \tag{9}$$

SVM has a property that it is completely determined by the training samples having the minimal distance to the plane and these samples are known as support vectors. The plane does not be changed by removing any or even all samples, other than support vectors. The hyperplane can thus be written as

$$f(x) = \text{sgn}(\sum_{i=1}^{n} c_i \alpha_i (x \cdot x_i) + b), \tag{10}$$

where the summation can be taken for i that having non-zero α_i. In (10), the term $(x \cdot x_i)$ measures the similarity between test sample x and training sample x_i that has been discussed in Section 2. If the samples of two classes overlap and they cannot be separated with any hyperplanes, the optimization problem can be formulated as

$$\text{Minimize } \frac{1}{2} w^T w + C \sum_{i=1}^{n} \xi_i \tag{11}$$

with constraint: $c_i (w^T x_i + b) \geq 1 - \xi_i, \quad i = 1,2,...,n$,

where ξ_i are slack variables.

After solving the above optimization problem with techniques similar to the separable case, optimal hyperplane (the soft margin hyperplane) can be determined. In summary, the training samples are used to determine the optimal hyperplane and class of test samples are predicted by the sign of it.

4 Threads Classification

This section describes the threads classification of a popular web forum. To avoid any advertising effects, the alias name R web forum is used in this paper. In the forum, threads are being tagged into different categories. There is an also category known as “General” category which stores posts not categorized by posters. A SVM classifier is constructed and used to classify the threads of the “General” category into five existing categories. Of course, there are also threads that do not fall into any of the five categories. Training samples are obtained from the threads of these categories which has the statistics shown in Table 1.

Table 1. Tatistic of the selected category

Category	No. of threads	No. of Posts	No. of Posters
Speakers	88	14,076	1,245
Integrated Amplifier	38	2,298	485
Tube Product	37	11,475	553
Cables	36	1,340	384
CD	34	2,489	412

4.1 Thread Vectors and Classifier

Thread-Poster-ID vectors and thread-Word-Content vectors were built first. Each vector's row had either Poster-ID or Word-Content to represent the features of a thread. In the experiment, five weighting scheme for vector component had been employed: TFIDF, $TFIDF^2$ (i.e. square of TFIDF), Term Frequency (TF), TF^2 and binary. From each of the category, 80% of the thread vectors were randomly selected

Table 2. Accuracy and std. dev. by using Poster-ID as vector features.

Weighting Scheme		Speakers	Integrated Amplifier	Tube Product	Cables	CD
TFIDF	Average	66.47	62.86	52.86	74.29	56.67
	Std. Dev.	14.18	27.26	16.96	26.95	15.28
$TFIDF^2$	Average	87.06	71.43	27.14	90.00	65.00
	Std. Dev.	14.60	34.40	18.57	30.00	36.09
TF	Average	69.41	58.57	51.43	54.29	50.00
	Std. Dev.	9.77	20.65	19.38	17.84	14.91
TF^2	Average	91.18	58.57	34.29	80.00	38.33
	Std. Dev.	5.42	34.08	23.21	33.93	27.94
Binary	Average	40.00	80.00	57.14	61.43	46.67
	Std. Dev.	9.41	13.09	27.11	25.59	26.67

Figure shown in percentage(%).

Table 3. Accuracy and std. dev. by using Word-Content as vector features

Weighting Scheme		Speakers	Integrated Amplifier	Tube Product	Cables	CD
TFIDF	Average	97.65	68.57	98.57	51.43	50.00
	Std. Dev.	2.88	29.83	4.29	20.40	29.81
$TFIDF^2$	Average	100.00	64.29	100.00	30.00	33.33
	Std. Dev.	0.00	44.38	0.00	43.07	40.14
TF	Average	88.82	68.57	84.29	77.14	80.00
	Std. Dev.	8.90	24.58	11.87	9.48	22.11
TF^2	Average	91.18	60.00	71.43	34.29	56.67
	Std. Dev.	8.00	38.76	24.74	23.21	27.08
Binary	Average	51.76	72.86	45.71	77.14	90.00
	Std. Dev.	10.12	25.11	27.70	24.07	17.00

Figure shown in percentage(%).

as training samples. The remaining threads were used as testing samples. It was repeated ten times for performance analysis. The results of using Poster-ID and Word-Content are shown in Table 2 and 3 respectively.

Referring to the statistics shown in Table 2 and 3, it is found that the classifier performance varies greatly against different weighting scheme and feature set. There is no single configuration of weighting scheme or feature set that outperform the other in all categories. However there is a trend that TFIDF is generally better in most case. The overall performance is relatively high and variation is steadier than the other weighting method.

4.2 Evaluation and Results

As each setup of the attribute weighting schemes has its own merit on different categories. A naïve approach of averaging the classifier scores is applied to yield a simple resultant value. A thread with its greatest resultant value for a category is classified to that category. Table 4 shows some sample threads originally placed in the "General" category are classified as class A (i.e. "Speakers" category).

Table 4. Sample classified results with Poster-ID and Word-Content as attributes

Thread subject	Feature Set of the Classifier	A	B	C	D	E
a question on speaker stand	Poster-ID	0.035	-0.002	-0.32	0.069	-0.204
	Word-Content	**0.775**	0.09	0.924	-0.293	-0.56
what cable is DENKO ?	Poster-ID	-0.638	0.707	-0.533	0.201	-0.376
	Word-Content	**0.769**	0.119	-0.043	0.117	-0.464
Focal,JM Lab VS Triangle	Poster-ID	0.576	-0.062	-0.351	0.01	-0.252
	Word-Content	**0.746**	-0.007	0.124	-0.275	-0.433
Any recommendation in audio equipment transportation?	Poster-ID	0.388	-0.01	-0.34	-0.063	-0.031
	Word-Content	**0.738**	0.018	-0.007	-0.41	-0.222

SVM Score of A: Speakers ; B: Integrated Amplifier; C: Tube Product; D: Cables; E: CD.

Table 5. Recall and precision of the classification against random retrieval and R tag searching

	Retrieval Method	A	B	C	D	E
Recall	Word-Content	93.48	97.10	67.07	68.52	95.38
	Poster-ID	73.09	N/A	81.72	58.32	N/A
	Random	8.06	5.14	5.08	5.02	4.91
	Tag Searching	63.77	55.07	45.12	33.33	52.31
Precision	Word-Content	91.68	73.27	61.52	82.20	70.73
	Poster-ID	73.09	N/A	77.02	73.24	N/A
	Random	8.06	4.03	4.79	6.30	3.80
	Tag Searching	100	100	97.30	100	100

Percentage(%) of A: Speakers ; B: Integrated Amplifier; C: Tube Product; D: Cables; E: CD.

Seeing the classified result as an IR system, the recall and precision of the classifier are analyzed against the taged searching provided by R and random retrieval which randomly pick threads from the whole pool.

In our retrieval analysis, only the first 50 returned threads plus the training/testing examples are evaluated. It is noticed that Word-Content features support the classifier for certain types of category to achieve a high recall and precision rate. While at some categories, Poster-ID classifier performs better than Word-Content classifier.

Tube Product in Hi-Fi society is a market niche where people are either interested or not interested in. This is an interest based category that requires special knowledge on it. Poster-ID as the features representing the behavior of the posters. Based on our experience and observation, posters are active in participating tube threads when they are interested only. This explains why Poster-ID classifier exceptionally perform better in tube product than the other categories.

The merit of Poster-ID classifier is less number of features are involved than words. It means less computation is required to classify a set of novel data. Also the processing of poster information is much simpler than message contents. Word stemming, stop words processing, segmentation (for non-English forums) are not necessary for posters.

However, Word-Content classifier has its advantages, it workw better in many cases. Words are the actual way to deliver meaning of a message. The barrier to achieving a good result may due to the incompleteness of corpus and/or unclear boundary of classes. Unsegmented senstences add difficulties to word analysis. Many of the Chinese segmentation algorithm was designed for Manda-Chinese/Written Chinese which do not have good performance on Canto-Chinese where English is mixed everywhere. Unfortunately, the medium of language of the forum was Canto-Chinese mixed with English. Although we have segmentated the senstence before analysis the classification however, some important corpus such as speakers and its chinese counterpart were not understanded by the classifier. Some jargon are even only understandable by human. Take an example: Model Number of a Product. This kind of word hide a number of feature behind, such as the brand/type of the product, which may be the important factor for making a classification decision. Also, two categories are sometimes hard to set a clear line to cut them into two pieces because they share too much features. This can explain why Word-Content classifier does not always work well in online communities.

5 Conclusion

In conclude, both Word-Content and Poster-ID classifier have their own merits based on different situations. The Poster-ID classified results can help reject a thread that is wrongly classified by Word-Content classifier and vice versa. On the other hand, a thread may belong to many categories as long as it is relevant. But for the sake of analysis, we assumed each thread can only belong to one category only. Therefore an approach combining the power of both type of classifier and analysis multiple tag of category will be a future work of this research.

References

1. Kleinberg, J. M.: Authoritative Sources in a Hyperlinked Environment. Journal of the ACM. Vol. 46 No. 5 (1999) 604-632.
2. Rogers, I., IPR Computing Ltd.: The Google Page Algorithm and How it Works. http://www.iprcom.com/papers/pagerank/.
3. Salton, G., McGill, M.J.: Introduction to Modern Retrieval. McGraw-Hill Book Company (1983).
4. Schölkopf B.: Statistical Learning and Kernel Method. MSR-TR 2000-23, Microsoft Research (2000)
5. Osuna, E., Freund, R., Girosi, F. : Support Vector Machine: Training and Applications. Technical Report AIM-1602, MIT AI. Lab (1996)
6. Salton, G., Wong, A., Yang, C.: Vector space model for automatic indexing. Communications of the ACM 18 (1975) 613-620
7. Salton, G., Buckley, C.: Term Weighting Approaches in Automatic Text Retrieval. Information Processing and Management, Vol. 24, No.5 (1988) 513-523
8. Vapnik, V.: The Nature of Statistical Learning Theory. Springer, N.Y. (1995)
9. Vapnik, V.: Statistical Learning Theory. Wiley, N.Y. (1998)
10. Omidvar, O., Dayhoff, J.: Neural Networks and Pattern Recognition. Academic Press (1998)

An Electronic Brokering Process for Truckload Freight

Kap Hwan Kim, Yong-Woon Choi, and Woo Jun Chung

Department of Industrial Engineering, Pusan National University,
Busan 609-735, Korea
kapkim@pusan.ac.kr, cantong@hanmail.net, wjjeong@lgcns.com

Abstract. This paper suggests a brokering process in which individual truckers participate in the process of matching trucks with freight. To maximize the total profit of trucks, a mathematical model is formulated for matching trucks with freight. It is shown that, by Lagrangean relaxation, the matching problem can be decomposed into a master problem and multiple sub-problems. This paper proposes a brokering process, based on the subgradient optimization technique, in which an optimality of individual truckers as well as the system optimality of truckers is satisfied. The effects of various practical considerations on the performance of the suggested brokering process are tested numerically and by using a simulation study.

1 Introduction

The brokering process in this study removes barriers between shippers and truckers and makes information on freight and trucks transparent to all participants. In this study, individual truckers can directly access information on freight, and each trucker can select a set of delivery orders from posted orders. Each trucker can consider his/her own personal preferences, constraints or conditions during the selection process of delivery orders. The brokering system requests shippers to propose maximum delivery charges they can pay for their freight, considering their experience or expectations on the availability of empty trucks. Because shippers and trucks at any place can utilize the brokerage system at lower brokering fees, more participation of shippers and truckers are expected, which result in the further reduction in the total transportation cost, compared to the case where small groups of shippers and truckers are independently involved in different brokerage companies.

In addition to the above advantages, the matches obtained by the brokering process must satisfy a system's optimality (in the sense that the total profit of trucks is maximized) as well as individual participant's optimality (in the sense that trucks are matched with freight based on each trucker's decision for maximizing his/her own profit). Although each trucker may be interested only in his/her own profit and not in the total profit of truckers, in the long run, decisions by individual truckers will result in the highest total expected profit to all truckers.

Related to the brokering process, several papers (Abrache et al.[1], Babin et al.[2], Kutanoglu and Wu[5]) on designing electronic market rules or auction processes have recently been published. Lagrangean relaxation technique has been suggested as one of the solution methodologies for solving combinatorial optimization problems

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 829 – 838, 2006.

(Reeves[7]). Kohl and Madsen[4] suggested an algorithm to solve the vehicle routing problem with time windows by using the subgradient optimization method and the bundle method, both of which are based on the Lagrangean relaxation model. Kutanoglu and Wu[5] proposed a resource scheduling method that uses an auctioning method based on a Lagrangean relaxation model for a manufacturing setting. Stewart and Golden[8] also suggested a heuristic vehicle routing algorithm based on the Lagrangean relaxation method. Zhao et al.[9] proposed a surrogated subgradient method for solving Lagrangean relaxation models.

The brokerage system in this paper comprises a brokerage center, shippers, and trucks. The procedure in which trucks are matched with delivery tasks is as follows:

1. A shipper sends a delivery application to a brokerage center through the internet, wireless communication, EDI, or a phone call. The application information includes the maximum delivery price the locations of pickup and delivery, the time windows for pickup, and other special requirements about delivery services.
2. The brokerage center monitors, receives, and maintains information about delivery orders and trucks and provides this information to shippers and truck drivers.
3. Trucks must be equipped with mobile computer terminals or TRS (trunked radio system) terminals. In some cases, truck drivers can access the web site of the brokerage system and download data about posted orders whose assignments are not yet fixed. The data about posted orders include the current minimum bidding price for each order and the truck ID that is most likely to be assigned to the order. Each truck driver selects the delivery orders and schedules the delivery travel that gives him/her the highest profit. Each truck driver informs the brokerage center of the list of selected orders. During the construction of the most profitable travel schedule, he/she must consider his/her local constraints and preferences.
4. The brokerage center collects bids from multiple trucks. Based on collected bids, the brokerage center adjusts the prices of orders if there are conflicts among different truck drivers or if there are no applicants for an order. The price of an order with more than one application will be decreased, while the price of an order with no application will be increased. However, the adjusted price cannot exceed the maximum price that the corresponding shipper initially submitted. The results of price modifications will be again informed truck drivers, or drivers may access the brokerage center to obtain information about the revised prices. Based on the revised prices, drivers can change their selection of delivery orders. This procedure is repeated until no conflicts exist or a time deadline for a commitment is reached.
5. At the deadline for a commitment, the brokerage center fixes the assignment and informs drivers and shippers of the results of the assignment.

The next section introduces a mathematical rationale of the brokerage process. Section three discusses the application of the subgradient optimization technique to the brokering process. The final section provides concluding remarks.

2 A Mathematical Formulation and a Subgradient Optimization Technique

This section proposes a mathematical model for maximizing the total profit of trucks, which is an objective from the system's perspective. However, it will be shown how

the model can be decomposed into multiple primal problems and a dual problem by Lagrangean relaxation and how each primal problem corresponds to the problem of maximizing each trucker's profit. That is, we will show that the system's optimality of truckers can be obtained by pursuing individual trucker's optimality. The following formulation assumes that each delivery order has a time window that allows for pickup of freight.

2.1 A Mathematical Formulation

The following are notations used in the formulation of the matching problem:

K : Set of trucks.
D^k : Node that represents the initial position of truck *k*.
D : Set of D^k.
H^k : Final destination of truck *k*, which may be the domicile of the truck.
H : Set of H^k.
L : Set of nodes that represent delivery tasks.
r_j : The price of delivery task *j* that a customer is willing to pay. This is the price that customer *j* initially posted on the application form. Note that $r_{H^K} = 0$.
v_j: The loaded travel cost for task *j*. Because it is assumed that all trucks are the same, this cost is also the same for all trucks.
x^k_{ij} : 1 if truck *k* performs task *j* after performing task *i*; 0, otherwise. This is the decision variable.
X^k : Vector whose elements are x^k_{ij}.
F^k : Set of X^ks that satisfy local constraints of truck *k*.
c^k_{ij} : Empty travel cost when truck *k* performs task *j* after performing task *i*. This is the empty travel cost from the delivery location of task *i* to the pickup location of task *j*. This cost includes expenses for fuel, labor, overhead for the truck, and other maintenance costs.
t_{ij} : Travel time from the delivery location of task *i* to the pickup location of task *j*.
t_i : Arrival time of a truck at the pickup location of task *i*.
$[a_i,b_i]$: Time window allowed for pickup of task *i*.
s_i : Service time for task *i*. This includes the pickup time at the pickup location of task *i*, the loaded travel time from the pickup location to the delivery location, and the drop-off time at the delivery location.
M : A large constant number.

The problem can be formulated as follows:

$$Max \sum_{k\in K} \sum_{j\in L\cup\{H^k\}} \sum_{i\in\{D^k\}\cup(L/\{j\})} (r_j - v_j - c^k_{ij}) x^k_{ij} \tag{1}$$

subject to

$$\sum_{k\in K} \sum_{i\in\{D^k\}\cup(L/\{j\})} x^k_{ij} \le 1, \text{ for all } j \in L \tag{2}$$

$$\sum_{j\in L\cup\{H^k\}} x^k_{D^k j} \le 1, \text{ for all } k \in K \tag{3}$$

$$\sum_{i\in\{D^k\}\cup L} x^k_{i\mathrm{H}^k} \le 1, \text{ for all } k \in K \tag{4}$$

$$\sum_{i\in\{D^k\}\cup L} x^k_{i\mathrm{H}^k} \le 1, \text{ for all } k \in K \tag{5}$$

$$a_i \le t_i \le b_i, \text{ for all } i \in L \tag{6}$$

$$\mathrm{t}_i + s_i + t_{ij} - t_j \le (1 - x^k_{ij})M, \text{ for all } j \in L \cup \{H^k\},\ i \in \{D^k\} \cup (L/\{j\}), \text{ and } k \in \mathrm{K} \tag{7}$$

$$X^k \in F^k, \text{ for all } k \in K \tag{8}$$

$$x^k_{ij} \in \{0,1\}, \text{ for all } j \in \{H^k\} \cup L,\ \ i \in \{D^k\} \cup (L/\{j\}), \text{and} \quad k \in \mathrm{K} \tag{9}$$

The purpose of the objective function (1) is to maximize the total profit. Constraint (2) implies that no delivery order can be assigned to more than one truck. Constraints (3) and (4) imply that a maximum of one arc can be selected as an exiting route or an entering route at the initial location or at the domicile. Flow conservation is satisfied by constraint (5). The arrival time of a truck at a pickup location is restricted by constraint (6). Constraint (7) represents the relationship between the time variables and the flow variable, and it prevents sub-tours in the solution. Local constraints of truck *k* can be expressed by constraint (8). Examples of local constraints are the time-window constraint at an initial location or a final domicile, the longest operation time allowed to a specific truck, etc. Note that there is no constraint to force all delivery orders to be included at least on one route because orders with excessively high prices may not be served by any truck. For problems with 10 trucks and 20 tasks, the number of variables and constraints were approximately 600 and 6000, respectively, excluding constraint (8).

This study attempts to solve problem (1) – (9) in a distributed manner. The Lagrangean relaxation technique is used to decompose the above problem. Let $r_j^{'} = r_j - v_j$. Then, by relaxing constraint (2), we can obtain the following subproblems (decomposed primal problems, or DPPs), which can be interpreted as the problem of a corresponding truck driver selecting delivery tasks and determining a schedule for performing the tasks in a way that maximizes his/her profit for given prices of tasks. The DPP for truck *k* can be written as follows:

(DPP) $$\text{Max} \sum_{j\in \mathrm{L}\cup\{\mathrm{H}^k\}} \sum_{i\in\{\mathrm{D}^k\}\cup(\mathrm{L}/\{\mathrm{j}\})} (\mathrm{r}_j^{'} - \lambda_j - \mathrm{c}^k_{ij})\mathrm{x}^k_{ij} \tag{10}$$

subject to constraints (3) - (9) where $\lambda_j \ge 0$ for all j and $r_H^{'} = \lambda_{H^K} = 0$.

2.2 Solving the Problem by the Subgradient Optimization Technique

The following notations are used to describe the algorithm: g_i is the subgradient of the node for task *i* and Z_{UB} and Z_{LB} is the upper bound and the lower bound of the objective function value. S is the step size for the Lagrangean multiplier and π is the parameter to adjust *s*. Then, a typical subgradient optimization procedure can be summarized as follows:

Step 1: (Initialize) Set $\lambda_i = 0$ for i=1, …, n; $\pi = 2$, N = 30.

Step 2: (Solve DPPs) For the given values of λ_i for i = 1, …, n, solve DPPs and let the resulting objective function value of (10) be Z_{UB}.

Step 3: (Resolve infeasibility) When more than one truck selects the same delivery order as the solution of a DPP, calculate for each corresponding truck the amount of decrease in profit resulting from canceling the selection of the order. Cancel selections for the trucks except that for the truck with the highest decrease in the profit. If the objective function value (1) for these revised solutions of DPPs is greater than the current lower bound, then update the lower bound.

Step 4: (Update multipliers) Calculate the values of the subgradients, as follows: $g_j = \sum_{k \in K} \sum_{i \in \{D^k\} \cup (L/\{j\})} x_{ij}^k - 1,\ j = 1,..,\mathrm{n}$. Update the step size and the multipliers by the following equations: $s_i = \dfrac{\pi(Z_{UB} - Z_{LB})}{\sum_{i=1}^{m} g_i^2}$, and $\lambda_i = \max(0, \lambda_i + s_i g_i)$

Step 5: If Z_{UB} was not updated during the last N consecutive iterations, then reduce π by half. If π is less than a specified value (0.005 in this study's numerical experiment) or if the duality gap (the difference between Z_{UB} and Z_{LB}) is less than a pre-specified value, the iterations are terminated. Otherwise, go to step 2.

3 Applying the Subgradient Optimization Technique to the Brokering Process

In the following, it was discussed how the subgradient method can be applied to the brokering process. The brokerage center adjusts the prices of delivery tasks, based on the number of bids submitted for each task. When a feasible assignment is not reached even after a reasonable amount of communication, the brokerage center makes a feasible solution based on the current solution and global information. The process is equivalent to the subgradient optimization procedure that modifies an infeasible solution to obtain a feasible solution. Individual truckers select profit-maximizing orders, based on prices of orders, locations of trucks, and other local constraints of trucks. This process is equivalent to solving a DPP in the subgradient optimization technique.

Notice that the above explanation on the solution procedure started from problem (1) – (9) whose objective is the system's optimality of maximizing the total profit of truckers. The previous section showed that problem (1) – (9) can be solved by solving DPPs – which are equivalent to problems whose objective is user's optimality of maximizing each trucker's profit – and adjusting the values of the Lagrangean multipliers, which is equivalent to the price adjustment by a brokerage center.

3.1 Performance of the Standard Subgradient Optimization Technique

Problems were solved to test the performance of the standard subgradient optimization technique. A program was developed and run on a Pentium II (333 MHz, 64 RAM) personal computer. The pickup and delivery locations were generated randomly on a rectangular area of 100×100 km. Initial prices for orders were also

generated randomly from a uniform distribution, U(100, 400), in units of dollars. Although, in practice, there must be some relationship between prices of orders and service times for performing the orders, the initial prices were randomly generated for providing orders with different profit margins, thus checking if truckers can select more profitable orders. Euclidean distance was used to measure the distance between two locations. When the difference between the upper bound and the lower bound came smaller than a specified small value, or when the number of iterations reached a pre-specified value, the iterations were stopped.

Two heuristic methods called GRASP (Greedy Randomized Adaptive Search Procedure) (Feo and Resende[3]) and NHS (Neighborhood Search) are applied to solve DPPs. NHS method used in this study is easier to implement than GRASP. Starting from an initial solution, NHS repeatedly moves to another feasible solution, which has the highest increase in the total profit, among neighborhood solutions. The neighborhood solutions were generated by inserting an order outside a route into the route, deleting an order from a route, and exchanging an order on a route with another order outside the route.

3.2 Analysis of Effects of Asynchronous Communication Between Trucks and the Brokerage Center

The decision process regarding matching must be robust enough to manage unexpected events. To improve the robustness of the decision process, a modified decision process is proposed in which the prices of delivery orders are updated whenever a decision by a trucker (a solution of a DPP) is transmitted to the brokerage center. That is, the brokerage center does not have to wait for the arrival of all decisions from participating trucks. This modification allows asynchronized communication between trucks and the brokerage center. The brokerage center may collect decisions from several truckers before it updates the prices, or it may wait until all decisions arrive, which will be the same as the synchronized communication.

The following experiment tested the effects of asynchronized communication on the quality of solutions. In the experiment, DPPs were solved sequentially by GRASP, and the values of the multipliers were updated whenever a DPP was solved. The solution procedure will be called the “GRASP-asynchronized method.” Let the subgradient optimization method, in which all the DPPs are solved before Lagrangean multipliers are updated, be called the “GRASP-synchronized method.” Considering that the amount of communication between the brokerage center and trucks must be limited and that the communication cost is not negligible, it is important to obtain a good solution at the lowest possible number of iterations. As the solution process progressed, the changes of the lower bound for the three methods were compared with each other. For a small number of iterations, the lower bound found by the GRASP-asynchronized method was higher than that created by GRASP-synchronized method. Thus, the GRASP-asynchronized method can be considered the better alternative than the GRASP-synchronized method when the number of communications must be restricted during the brokering process.

A numerical experiment compared the performance of the GRASP-asynchronized method with that of the GRASP-synchronized method. The problem set (problems 11-20) with 7 trucks and 15 tasks and the problem set (problems 21 – 30) with 20

trucks and 50 tasks were solved by both methods. The maximum number of iterations was limited to 210. The results showed that the objective values found by the GRASP-asynchronized method were higher than those found by the GRASP-synchronized method by 19% on average, which is a significant improvement. Also, the average computational time required by the GRASP-asynchronized method was 98 seconds, which is only 26% of that needed by the GRASP-synchronized method

In the synchronized method, the number of communications between each truck and the brokerage center was 210 during 210 iterations. However, in the asynchronized method, it was about 10, because, in the asynchronized method, only one DPP was solved during each iteration and the results are transmitted to the brokerage center.

3.3 Comparison Between GRASP-Asynchronized and NHS-Asynchronized Methods

Four sets of problems of different sizes were constructed: the problem set (problems 31-60) with 10 trucks and 40 delivery orders, the problem set (problems 61-90) with 30 trucks and 120 orders, the problem set (problems 91-120) with 50 trucks and 200 orders, and the problem set (problems 121-150) with 70 trucks and 280 orders. Each truck has the maximum travel distance of 5,000 during the planning horizon.

An experiment was conducted to compare the results of the GRASP-asynchronized algorithm with that of the NHS-asynchronized algorithm for the above four sets of problems. When the average of 30 problems was compared, the objective values found by the NHS-asynchronized algorithm were approximately 91% of those by the GRASP-asynchronized algorithm. The computational time needed by the NHS-asynchronized algorithm was 0.3 – 4.0 % of that needed by the GRASP-asynchronized algorithm, and the percentage tended to decrease as the size of problems increased. Because computational time was considered to be critical to implementing the algorithm in real time, it was concluded that the NHS-asynchronized algorithm is appropriate for practical uses.

Also, it was found that when the NHS- asynchronized algorithm was used, although the initial objective value was low, the objective value approached its maximum within several seconds, while the GRASP-asynchronized algorithm required more than 2,000 seconds to stabilize the objective value. For other problems, similar differences were observed.

3.4 Performance Evaluation of the Brokering Process in This by a Simulation

A simulation program using JAVA language was developed to test the performance and applicability of the brokering process of this study (BROK) in a dynamic situation where candidate trucks and new delivery orders arrive at the system continuously and randomly. First, at the beginning of the simulation, trucks were randomly generated on a square of 100×100. And then, delivery orders were also generated randomly on the square, however, the generation times of the orders were selected randomly between 0 and 100. The lower and upper bounds of the pickup time window for an order were also randomly selected. The initial prices of orders were also selected randomly.

A conventional dispatching process (CONV) was compared with BROK. It was assumed that CONV does not use the brokering system but all the information on delivery orders and trucks are available to all the shippers and truckers. CONV matches trucks and delivery orders as follows: When a trucker completes a delivery task, he/she selects the order that gives him/her the highest profit. When a shipper wants to issue a delivery order, he/she selects the truck with the least delivery cost. It was also assumed that the shipper pays the initial price for the delivery order to the corresponding trucker.

In BROK, the NHS-asynchronized algorithm is triggered whenever a delivery order arrives at the system. Thus, the assignment of orders to trucks may change by an arrival of a new order. The first order in the schedule of a truck is finally assigned to the truck only when the truck completes an order. In BROK, a truck has a temporary schedule for the future travel, while, in CONV, a truck can have at most one assigned delivery order.

Four different situations were simulated. The four situations had 20, 30, 40, and 50 trucks, and 200, 300, 400, and 500 delivery orders, respectively. For each situation, the simulation was run 20 times. It was found that a higher ratio of orders was served by trucks in BROK compared to that in CONV. And, the transportation cost per order was lower in BROK than in CONV. The average price per order paid by shippers was also significantly lower in BROK than in CONV.

The profit per order and the total profit earned by trucks were lower in BROK than in CONV, which is due to the reduced price resulting from competition for orders among truckers. If some portion of the reduced price is redistributed to truck drivers, then both sides (truck drivers and shippers) may be better off in BROK. For example, suppose that 50% of the price reduction in BROK is reserved for the redistribution to truckers, which means that the price of the delivery is discounted by only 50% of the price difference between the initial price and the final price. It was found that the total profit of all the trucks increased significantly compared with that of CONV, while a significant price reduction is still provided to the shippers.

3.5 Issues for Implementation

As shown in the previous subsection, regarding shippers, because the final price of delivery task *j* is expressed as r_j-λ_j, the final price of all tasks must be less than or equal to the original prices suggested by shippers. Thus, shippers must have sufficient incentive to utilize the brokerage system for their freight.

Regarding truckers, because the brokering process already guarantees individual trucker's optimality, there is no other opportunity for a trucker to earn more money by changing his/her decision. If a trucker decides that the average price of tasks is lower than what he/she expects, he/she may leave the system to find tasks outside the system. If the number of trucks decreases, then the average price of tasks will remain at a certain profitable level. However, in the introduction stage of the system, truckers may complain that their profit per order is reduced, and this may be true. Considering that more shippers will become customers of the brokerage system and thus truckers may carry more freight than before, the total profit of a truck may increase in the long run. If the increased amount of freight is not enough of an incentive for truckers to participate in the brokerage system, then the difference between the initial price (r_j)

and the final price (r_j - λ_j) of a freight may be shared between the shipper and the trucker of the freight as shown in the previous subsection.

Suppose that all trucks are owned by a single carrier. Then, the decisions about matching for all the trucks are made by a centralized matching department or personnel. The delivery charge for freight is usually determined by negotiation between the trucking company and the shipper of the freight. Thus, it can be assumed that shippers must pay the prices that they initially proposed. However, a fixed salary is usually paid to drivers on a monthly basis. Thus, the trucking company takes all the savings resulting from the optimized matching.

4 Conclusion

This paper introduces a brokerage system that assigns delivery tasks to trucks that are operated as independent profit-makers. The matching method in this study is distributed in that the decisions about matching are made through communications between a brokerage center and trucks. It has been shown that the matching problem can be decomposed into multiple sub-problems that processors in trucks can solve by using local information. Also, it was shown how the subgradient optimization method can be utilized as a theoretical basis for the distributed matching procedure.

The asynchronized procedure based on the neighborhood search (NHS-asynchronized method) was proposed as the most promising algorithmic basis for the electronic brokering process. Through a simulation study, it was also found that the suggested brokering process significantly outperforms a conventional dispatching process. Also, it was shown that the total reduction in prices resulting from the suggested brokering process exceeds the total reduction in profits of truckers significantly so that both shippers and truckers can be better off by sharing benefits of shippers with truckers.

Acknowledgments

This work was supported by the Regional Research Centers Program (Research Center for Logistics Information Technology), granted by the Korean Ministry of Education & Human Resources Development.

References

1. Abrache, J., Crainic, T. G., and Gendreau, M.: Design Issues for Combinatorial Auctions. CRT-2002-2, Centre de recherché sur les transports, University of Montreal. (2002)
2. Babin. G., Crainic, T. G., Gendreau, M., Keller, R. K., Kropf, P., and Robert, J. :Towards Electronic Marketplaces: A Progress Report. CRT-2001-33, Centre de recherché sur les transports, University of Montreal. (2001)
3. Feo, T. A. and Resende, M. G. C.: Greedy randomized adaptive search procedures. J. of Global Optimization, Vol. 6, pp.109-133. (1995)

4. Kohl, N. and Madsen, O. B. G.: An Optimization Algorithm for the Vehicle Routing Problem with Time Windows Based on Lagrangean Relaxation. Technical Report, Institute of Mathematical Modeling, The Technical University of Denmark (http://www.imm.dtu.dk/ /documents/ftp/tr95/tr95.html).
5. Kutanoglu, E. and Wu, S. D.: On combinatorial auction and Lagrangean relaxation for distributed resource scheduling. IMSE Technical Report 97T-012, Lehigh University, (http://lehigh.edu/ ~sdw1/erhan1). (1997)
6. Oren, M. E. and Williams, A. C.: On competitive bidding. Operational Research, Vol. 23, No. 6, pp. 1072-1079. (1975)
7. Reeves, C.: Modern Heuristic Techniques for Combinatorial Problems. McGraw-Hill International, UK
8. Stewart, Jr,. W. R. and Golden, B. L. : A Lagrangean relaxation heuristic for vehicle routing. European Journal of Operational Research, Vol. 15, pp. 84-88. (1984)
9. Zhao, X., Luh, P. B., and Wang, J.: The surrogate gradient algorithm for Lagrangian relaxation. Journal of Optimization Theory and Applications, 100(3) (http://www.engr.uconn.edu/ msl/p7.htm). (1999)

A Fuzzy Integral Method of Applying Support Vector Machine for Multi-class Problem

Yanning Zhang[1], Hejin Yuan[1], Jin Pan[2], Ying Li[1], Runping Xi[1], and Lan Yao[1]

[1] School of Computer Science, Northwestern Polytechnical University
710072 Xi'an, China
[2] Lab of Network Security and Countermeasure, Xi'an Communication Institute
710106 Xi'an, China

Abstract. This paper proposed a novel method of applying support vector machine for multi-class problem based on fuzzy integral. Firstly, the fuzzy measure of each binary classifier is constructed based on its classification accuracy during training and its agreement degrees to other support vector machines. Then the testing instances are classified by calculating the fuzzy integral between the fuzzy measures and the outputs of the binary support vector machines. The experiment results on iris and glass datasets from UCI machine learning repository and real plane dataset show that the new method is effective. And the experiment results ulteriorly indicate that the method with Choquet fuzzy integral has better performance than that with Sugeno integral.

1 Introduction

Support vector machine (SVM) [1] [2] is one of the most important achievements of statistical learning theory. Classifier constructed by this method can not only classify the samples but also has a maximal margin between the two classes. The details about recent developments of SVM can be found in [3] [4]. SVM was originally put forward for binary classification problem. While multi-class problem has more significance in reality, so how to apply support vector machine for these problems is now a hot topic. Generally speaking, there are two kinds of methods. One is constructing multiple binary support vector machines and classifying samples by combination of them, e.g. one-against-one method [5], one-against-rest method, decision directed acyclic graph support vector machine(DDAGSVM)[6][7], binary tree support vector machine [8], etc; the other is extending the standard support vector machine model for multi-class problem directly through modifying the object function. As to the second method, the training speed is very slow since there are too many parameters needs to be optimized. Moreover, its performance is not very ideal in fact [9]. The voting method adopted by one-against-one method is not very proper, because its voting value is either one or zero no matter what the real output of the classifier is. In addition, the classifier's difference is also not considered. DDAGSVM adopts the thought of direct acyclic graph during classification, so it needn't calculating all of the binary classifier's outputs. However, this method does not consider the current classification error's influence on the subsequent operation and its performance is dependent on the

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 839 – 846, 2006.

specific direct acyclic graph. One-against-rest method turns k class problem into k binary classification problems by taking one class samples as "+" and the others as "-". Its shortcoming is it requests every sample's participation during training. And its performance is no better than one-against-one method in practice. For binary tree support vector machines, how to build the tree is a difficult problem and there is no very effective method till now. By applying some ideas from the Error Correcting Output Codes techniques [10], the authors propose a combinatorial distribution of the training patterns of different classes in two classes in order to maximize the Hamming distance between classes. More sophisticated combinations can be made if machine's outputs are considered probabilities [11] or SVCs are adapted to produce posterior class probabilities as pointed in [12]. Finally, in the last few years, several other approaches have been developed considering all the classes at once multi-class SVCs [13] [14].

During research, we found if we took each binary support vector machine as a sensor, then the procedure combining them to solve k class problem can be treated as an information fusion procedure on the decision level. This discovery gives us some new inspiration with regard to applying support vector machine for multi-class problems. There are many useful algorithms for information fusion, such as Bayesian theory, Dempster-Shafter theory, neural networks and fuzzy integral. Among them, fuzzy integral is a simple and effective method because it needn't any prior-information and can prettily synthesize the objective evidence and the subjective decision. Based on all analysis above, this paper proposed a novel method applying support vector machine for multi-class problem with fuzzy integral. This method is similar to one-against-one method, while it adopts fuzzy integral to replace the voting method.

The rest of this paper is organized as follows: section II provides a brief introduction to support vector machine; some basic conception and properties about fuzzy measure and fuzzy integral are described in section III; section IV shows the details about the fuzzy measure construction and the method using fuzzy integral to solve multi-class problem with support vector machine; section V gives the experiment results and analysis of our method on some datasets; finally, we draw a conclusion in section VI.

2 Support Vector Machine

Support vector machine was originally formulated to construct binary classifiers from a given set of labeled training samples defined by $\{x_i, y_i\} \in \mathrm{R}^{\mathrm{N}} \times \{\pm 1\}$,where $i = 1,2,\cdots,n$. Support vector machine obtains its higher generalization ability by maximizing the margin between the nearest samples of the two classes. The decision function of support vector machine is as following:

$$f(x) = \mathrm{sgn}(w^* x + b) = \mathrm{sgn}\left(\sum_{i=1}^{n} a_i^* y_i K(x_i, x) + b^* \right) \tag{1}$$

where $f(x)=0$ corresponds to the optimal hyperplane. If $f(x)>0$, x is assigned to the positive class, otherwise it is assigned to the negative class. Here, $K(\cdot,\cdot)$ is the kernel function. a_i^* are called Lagrange multipliers and are calculated when solving the following quadratic programming problem:

$$\begin{aligned} &Max(Q(a))=\sum_{i=1}^{n}a_i-\frac{1}{2}\sum_{i,j=1}^{n}a_i a_j y_i y_j K(x_i,x_j) \\ &st.\ \sum_{i=1}^{n}a_i y_i=0 \qquad 0\le a_i\le C \quad (i=1,2,\cdots,n) \end{aligned} \tag{2}$$

The vectors with nonzero coefficients are called support vectors. C is a regularization parameter which controls the tradeoff between complexity of the machine and the number of nonseparable points.

In standard support vector machine model, w^*x+b reflects the distance between x and the decision boundary. The method deciding the class label of x by symbolizing w^*x+b is too rough, because it doesn't reflect the real degree that x belongs to this class. Too much information is lost during this procedure. So, here we use (3) to compute the fuzzy membership degree when x is determined to be positive sample.

$$f_+(x)=\begin{cases}1 & \text{if} \quad w*x+b\ge 1\\ w*x+b & \text{else}\end{cases} \tag{3}$$

Similarly when it is determined to be "-":

$$f_-(x)=\begin{cases}1 & \text{if} \quad w*x+b\ge 1\\ |w*x+b| & \text{else}\end{cases} \tag{4}$$

3 Fuzzy Measure and Fuzzy Integral

Fuzzy measure and fuzzy integral were firstly introduced by Sugeno [15]. In mathematical terms, a fuzzy measure is a set of functions with monotonic property but not always additivity. Fuzzy integral is a fusion operator which combines nonlinearly, objective evidences in the form of expert response with respect to a subjective evaluation of their performance given in the form of fuzzy measures.

3.1 Fuzzy Measure

A commonly used fuzzy measure is g_λ fuzzy measure described in definition 1.

Definition 1: if $\lambda\in(-1.+\infty), g_\lambda:\Omega\to[0,1]$ has the following properties, it is called g_λ fuzzy measure:

1) $g_\lambda(u)=1$;

2) if $A,B\in\Omega$ and $A\cap B=\phi$, then $g_\lambda(A\cap B)=g_\lambda(A)+g_\lambda(B)+\lambda g_\lambda(A)g_\lambda(B)$;

3) g_λ is continuous.

Definition 2: Suppose $X=\{x_1,x_2,\cdots,x_n\}$ is a finite set, and $g_i=g_\lambda(\{x_i\})$, then the mapping $g_\lambda:x_i\to g$ is called a fuzzy density function.

For g_λ fuzzy measure, it has the property described by theorem1.

Theorem 1: Suppose $X=\{x_1,x_2,\cdots,x_n\}$ is a finite set, and $\{g_i\}(i=1,2,\cdots,n)$ is the fuzzy density function of it, then there exist the only one $\lambda(\lambda\neq 0)$ which fulfills the equation $\prod_{i=1}^{n}(1+\lambda g_i)=\lambda+1$.

From above, we know if we want to calculate the fuzzy measure of a given set, we only need to know its fuzzy density function and then use theorem 1 to solve out λ, and finally use the second item in definition 1 to generate g_λ fuzzy measure.

3.2 Fuzzy Integral

Fuzzy integral is a generalized inner product between fuzzy measure and fuzzy membership function. There are many different kinds of fuzzy integrals. Here we only simply introduce Sugeno and Choquet fuzzy integral [16].

Definition 3: Suppose $\{U,\Omega,g\}$ be a fuzzy measure space and $h:\Omega\to[0,1]$ be a measurable function. Sugeno fuzzy integral of the function f with fuzzy measure g over $A\in\Omega$ is defined as:

$$(S)\int_A f(x)\circ g(\cdot)=\underset{\lambda\in[0,1]}{Sup}(\lambda\wedge g(A\cap h_\lambda)) \tag{5}$$

Where $h_\lambda=\{u\mid h(u)\geq\lambda\},0\leq\lambda\leq 1$ and $g(\cdot)$ is the measure of the set A. Specially, if $A=U=\{u_1,u_2,\cdots,u_n\}$, $h(u_1)\geq h(u_2)\geq\cdots\geq h(u_n)$, $A_i=\{u_1,u_2,\cdots,u_i\}$ and the fuzzy measure is g_λ measure, Sugeno fuzzy integral can be calculated as follows:

$$(S)\int_A f(x)\circ g(\cdot)=\bigvee_{i=1}^{n}(h(u_i)\wedge g_\lambda(A_i)) \tag{6}$$

Definition 4: Choquet fuzzy integral of the function f with fuzzy measure g over A is given as follow:

$$(C)\int_A f(x)\circ g(\cdot)=\int_0^1 g(h_\alpha)d\alpha \tag{7}$$

Here $h_\alpha=\{u\mid h(u)\geq\alpha\},0\leq\alpha\leq 1$. If $A=U=\{u_1,u_2,\cdots,u_n\}$, $h(u_1)\leq h(u_2)\leq\cdots\leq h(u_n)$ and the fuzzy measure is g_λ fuzzy measure, then Choquet fuzzy integral can be calculated as:

$$(C)\int_A f(x)\circ g(\cdot)=\sum_{i=1}^{n}[h(u_i)-h(u_{i-1})]g_\lambda(A_i)=\sum_{i=1}^{n}h(u_i)[g_\lambda(A_i)-g_\lambda(A_{i-1})] \tag{8}$$

Where $h(x_0)=0$, $g_\lambda(A_{n+1})=0$ and $A_i=\{u_1,u_2,\cdots,u_i\}$.

4 Applying SVM for Multi-class Problem with Fuzzy Integral

If we take each binary support vector machine as a sensor, the procedure combining them to solve k class problem can be considered as a decision information fusion process. Based on this idea, this paper presented a novel method to solve multi-class classification problem by support vector machines with fuzzy integral. Like one-against-one method, it also needs C_k^2 standard support vector machines. The difference is that it uses fuzzy integral to synthesize the information of the binary classifiers rather than voting method.

The critical point for this method is how to construct an appropriate fuzzy measure to represent the binary classifier's importance. We think such factors should be considered: the uncertainty of the classifier's output, the agreement degree among the classifiers for the same conclusion and the reliability of the classifiers. Since the first element has been considered in section 2, here we only put attention to the last two factors.

Using Deng's work [17] as a reference, we adopt the following method to represent fuzzy measure of the binary support vector machine. For instance x, there are $k-1$ binary support vector machines denoted as $C_{i,1}, C_{i,2}, \cdots, C_{i,i-1}, C_{i,i+1}, \cdots, C_{ik}$ $(i=1,2,\cdots,k)$ to decide the degree which it belongs to class i. And their outputs are denoted as $Y_i=\{y_{i1}/_{Ci,1}, y_{i2}/_{Ci2},\cdots, y_{i,i-1}/C_{i,i-1}, y_{i,i+1}/C_{i,i+1},\cdots, y_{i,k}/C_{ik}\}$. Then the procedure using fuzzy integral to fusion the information is actually a fuzzy transformation of Y_i. Suppose $h_{i,m}(x), h_{i,n}(x)$ are the outputs of classifier $C_{i,m}$ and $C_{i,n}$ for instance x. The agreement degree between $C_{i,m}$ and $C_{i,n}$ is defined as $S_{m,n}=\dfrac{\min\{h_{i,m}(x),h_{i,n}(x)\}}{\max\{h_{i,m}(x),h_{i,n}(x)\}}$. For class i, we can calculate the agreement degree between each pairs from $C_{i,1}, C_{i,2},\cdots, C_{i,i-1}, C_{i,i+1},\cdots, C_{ik}$ and get a agreement degree matrix as follow:

$$S_i=\begin{bmatrix} 1 & s_{1,2} & \cdots & s_{1,i-1} & s_{1,i+1} & \cdots & s_{1,k} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ s_{i-1,1} & s_{i-1,2} & \cdots & 1 & s_{i-1,i+1} & \cdots & s_{i-1,k} \\ s_{i+1,1} & s_{i+1,2} & \cdots & s_{i+1,i-1} & 1 & \cdots & s_{i+1,k} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ s_{k,1} & s_{k,2} & \cdots & s_{k,i-1} & s_{k,i+1} & \cdots & s_{1,k} \end{bmatrix}$$

This matrix intuitively gives us the agreement degree of each classifier for instance x. The average agreement degree of $C_{i,j}$ is defined as:

$$M(C_{i,j}) = \frac{1}{k-1} \sum_{t=1,t\neq i,j}^{k} s_{j,t} \tag{9}$$

In order to consider the reliability of the classifier, we take the classification accuracy as a reference when constructing the fuzzy measure. Then the final fuzzy measure of classifier $C_{i,j}$ is defined as follow:

$$g_{i,j} = \alpha M(C_{i,j}) + (1-\alpha) r_{i,j} \tag{10}$$

Where $\alpha \in [0,1]$ is a coefficient, and $r_{i,j} \in [0,1]$ represent the right classification rate of classifier $C_{i,j}$ during training.

With fuzzy measure, the computation of fuzzy integral will be s simple procedure. The details steps using support vector machine to solve multi-class problem with fuzzy integral are described as follows:

1) For the given sample dataset, train C_k^2 standard support vector machines as in one-against-one method and statistic their right classification rates;

2) For an arbitrary instance x, use $C_{i,1}, C_{i,2}, \cdots, C_{i,i-1}, C_{i,i+1}, \cdots, C_{ik}$ to calculate the degree of which belongs to class i, and the results are denoted as $h_{i,1}(x), h_{i,2}(x), \cdots, h_{i,i-1}(x), h_{i,i+1}(x), \cdots, h_{i,k}(x)$;

3) Calculate the agreement degree of each pair of the classifiers and generating each classifier's fuzzy measure with formula (11) ; the finally result is denoted as $g_{i,1}(x), g_{i,2}(x), \cdots, g_{i,i-1}(x), g_{i,i+1}(x), \cdots, g_{i,k}(x)$;

4) Computing Sugeno integral or Choquet integral of $h_{i,1}(x), h_{i,2}(x), \cdots, h_{i,k}(x)$ and $g_{i,1}(x), g_{i,2}(x), \cdots, g_{i,k}(x)$, then the output $\mu_i(x)$ represents the degree that x belongs to class i;

5) The final classification result for instance x is $c = \arg\max_i \mu_i(x)$, $i = 1,2,\cdots,k$.

5 Experiments and Analysis

In order to verify the performance of our method, we carry out two groups of experiments. One is on the Iris and glass datasets from UCI machine learning repository [18] and the other is from some real dataset of plane. For each dataset, one half of them are randomly selected for training and the rest of them are used as testing dataset. We also test one-against-one method and one-against-rest method in comparison of our method with Sugeno and Choquet fuzzy integral respectively. The results are shown in table 1.

Table 1. Statistical result for the different algorithms

Dataset			Right classification rates of the algorithms (%)			
Name	Class num	Sample num and dimension	One-against-one	One-against-rest	Sugeno fuzzy Integral	Choquet fuzzy integral
Iris	3	150,4	93.33	92.00	92.00	96.00
Glass	6	214,9	77.14	71.43	72.89	78.09
Plane	3	1084,64	94.30	89.65	90.10	95.41

From table 1, we can find for each dataset, Choquet fuzzy integral has the best performance while one-against-rest is the worst of them. One-against-one method is a little bit better than Sugeno fuzzy integral. We think these experiment results are reasonable. From definition 3, we know Sugeno fuzzy integral is actually a manner of finding the best match between the objective evidence and subjective expectation. Its most obvious characteristic is strengthening the principal factor's effect while neglecting the subordinate factors completely. This property may not be very reasonable for combining support vector machines for multi-class problems, because many support vector machines' functions are ignored. While one-against-one method consider each binary classifier's output, so it has better performance than Sugeno fuzzy integral. Choquet fuzzy integral can be considered as the weighted sum of $h_i(x)$ and its weight is dependent on the order of $h_i(x)$. So Choquet fuzzy integral is actually a nonlinear synthesis between the classifier's output and importance. In Choquet fuzzy integral method, each binary classifier has its own weight which represents its importance and its output has more flexibility, so it is not difficult to understand why it has better performance than one-against-one method.

6 Conclusion

This paper presented a novel method of applying support vector machine for multi-class classification problem with fuzzy integral. The experiments on standard dataset and real dataset show that our method is more effective. Comparing to other methods, the superiority of fuzzy integral is that through fuzzy measure the classifiers' importance difference is considered appropriately in the final decision. And the fuzzy measure of the classifier is not fixed. It is dependent on the specific sample, so it has a good adaptive property.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (NSFC) under Grants 60472072, the Specialized Research Foundation for the Doctoral

Program of Higher Education under Grant 20040699034 the Aeronautical Science Foundation of China under Grant 04I50370 and the Natural Science Foundation of Shaan'xi Province.

References

1. Vapnik. The Nature of Statistical Learning Theory. NewYork: Springer-Verlag ,1995.
2. Cortesc, Vapnik. Supporter Vector Networks. Machine Learning, 1995,20: 273-297.
3. Kecman, V. Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models. Cambridge, Mass. MIT Press, 2001.
4. Wang, L.P. (Ed.): Support Vector Machines: Theory and Application. Springer, Berlin Heidelberg New York,2005.
5. Botton L, Cortes C, Denker J, et al. Comparison of Classifier methods: A Case Study in Handwritten Digit Recognition. Proc of the Int Conf on Pattern Recognition. Jerusalem,1994:77-87.
6. Platt J, Cristianini N, Shawe-Taylor J. Large Margin DAG's for Multiclass Classification. Advances in Neural Information Processing Systems. Cambridge, MA:MIT Press,2000:547-553.
7. Takahashi F, Abe S. Decision-Tree-Based Multiclass Support Vector Machines. Proc of the 9th Int Conf on Neural Information Processing. Singapore.2002:1418-1422.
8. Sungmoon C, Sang H.O, Soo-Young L. Support Vector Machines with Binary Tree Architecture for Multi-Class Classification. Neural Information Processing-Letters and Reviews,2004,2(3):47-51.
9. Weston J, Watkins C. Mult-class Support Vector Machines. Technical Report. CSD-TR-98-04. Department of Computer Science. Royal Holloway University of London, England, May 1998.
10. T.G. Dietterich, G. Bakiri. Solving Multiclass Learning Problems via Error-correcting Output Codes. Journal of Artificial Intelligence,1995,263-286.
11. M. Moreira, E. Mayoraz. Improved Pairwise Coupling Classification with Correcting Classifiers. Proceedings of the ECML-98, Chemnitz,Germany,1998:160-171.
12. J. Platt. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. Advances in Large Margin Classifiers, MIT Press, Cambridge, 1999.
13. Y.Guermeur, A. Elisseeff, H. Paugam-Moisy. A New Multi-class SVM Based on a Uniform Convergence Result. Proceedings of the IJCNN-00,Como,Italy,2000
14. Cecilio Angulo, Xavier Parra, Andreu Catala. K-SVCR A Support Vector Machine for Multi-class Classification. Neurocomputing,2003:55-77.
15. Sugeno M. Fuzzy Measure and Fuzzy Integrals—A Survey. Fuzzy Automation and Decision Processes. New York, North-Holland,1977:89-102.
16. Murofushi T, Sugeno M. A Theory of Fuzzy Measures: Representations, the Choquet Integral, and Null Sets. Journal of Math Analysis and Applications,1991,32(2):532-549.
17. Deng Yong , Shiwen Kang, Zhao Chunjiang. A Method to Aggregate Fuzzy Sensor Data. Journal of Shanghai Jiaotong University, 2003,37(1):1-4.
18. http://www.ics.uci.edu/~mlearn/MLRepository.html

A Low-Power CMOS Analog Neuro-fuzzy Chip

Wang Wei-zhi[1] and Jin Dong-ming[2]

[1,2] Institute of Microelectronics, Tsinghua University, 100084,
Beijing, China
wangwz02@mails.tsinghua.edu.cn,
jdm-ime@tsinghua.edu.cn

Abstract. This paper proposes a low-power high-speed CMOS analog two-input/one-output neuro-fuzzy chip, the basic blocks of which are Gaussian membership function circuit, minimization circuit, and a centroid algorithm defuzzification circuit. This chip has been fabricated in 0.6-μm mixed-signal CMOS technology. Experiment and HSPICE simulation results show that because the compact structures and high-precision of the blocks the system not only consumes little power but also can be directly tuned by the weights calculated through the software such as Matlab.

1 Introduction

Neuro-fuzzy systems have been used successfully in various perplexing control tasks because of their straightforward architectures and no requirement of accurate mathematic models. Implementations of large control algorithms with many variables are usually carried out by software, such as mechanical or thermal processes, which typically operate below K FLIPS rate (FLIPS stands for fuzzy logic inferences per second) [1]. It is not fast enough for many high-speed control problems such as automotive engines control and power systems [2]. To manage delays of a few milliseconds and down to the microseconds and even nanoseconds range, ASICs are required [1]. Digital ASICs [2] can handle complex tasks with high precision, but they need additional analog circuits to build the interface with sensors and actuators. Analog ASICs [1,3] are worth considering, providing parallel processing with the speed limited only by the delay of signals through the network so that it has intrinsically higher speed and lower power consumption than their digital counterparts. Furthermore, because of straight interface to the plant and faster operation analog circuits can get better results in terms of less overshoot, smaller settling time, oscillations, ripple voltages, or currents [1].

In this paper, a high-speed programmable two-input/one-output analog CMOS neuro-fuzzy chip is proposed. Because of the compact structures as well as the precision of the analog blocks, this chip consumes the least power compared with the recent contributions and can be directly tuned by the weights calculated through the software such as Matlab. The chip has been fabricated in 0.6-μm mixed-signal CMOS technology. The operation speed of inference in the designed neuro-fuzzy system is 3 MFLIPS with the power-consumption of only 3.6mW.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 847 – 850, 2006.

2 Analog Function Blocks for Neuro-fuzzy System

2.1 Gaussian Membership Function Circuit

Gaussian shape membership function can obtain smoother transient response and facilitate the analog implementations. Differential pairs can be used to generate a current increasing from zero to the tail current of the circuit. And such two increasing currents can form the Gaussian-like waveform [3]. The compact circuit of programmable Gaussian membership is shown on the left in Fig. 1.

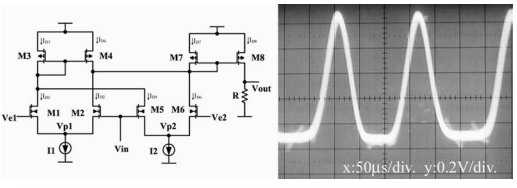


Fig. 1. Gaussian membership function circuit

From the discussion in [3], if the reference voltage Ve1 is set below Ve2, I_{D2} will be always lager than I_{D5} so as to rise to I ahead of I_{D5}. Therefore, the waveform of I_{D8}, i.e. the difference between the two drain currents, is alike as Gaussian shape, and so is the output voltage Vout. The test result is shown on the right in Fig. 1 with an input triangle signal changing from 0 to 1V. The mismatches between M2 and M5, I1 and I2 will make the Gaussian waveform asymmetric. And the mismatch between the two tail current sources also introduces output offset.

2.2 Minimization Circuit

The proposed minimization circuit uses PMOS differential input transistors M1 and M2 (or M6 and M7) in Fig. 2 (a) to remove the DC level shift circuits, which were used to ensure proper gate-biasing of transistor M5 and M10 over a wide range of output voltages, and change the input range to fit the output range of the former membership function blocks as well [3]. M5 and M10 form the based voltage minimization circuit. M1-M4 forms the PMOS input stage amplifier so that the input voltage, i.e. the output of the former membership function circuit, can be lower to V_{SS}. With the feedback, M1-M5 or M6-M10 constructs a unity gain buffer, making the output voltage just equal to the minimization input voltage. Fig. 2 (c) and (b) show the test output voltage waveforms of the minimization circuit compared with the ideal outputs.

Mismatch between M1 and M3, or M2 and M4, will introduce offset and increase the error. Each unity gain buffer is a two-stage amplifier, and each node in the circuit introduces a pole that is associated with the conductance and parasitic capacitance seen from the node. However, in this case the output pole is high frequency pole because the next stage input capacitance, the load of the circuit, is of little value. So no compensation capacitance is needed to guarantee the stability.

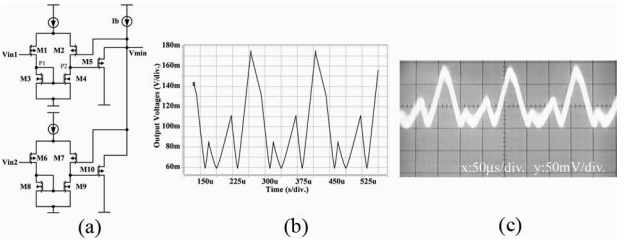


Fig. 2. Two inputs minimization circuit

2.3 Centroid Defuzzification Circuit

Parallel conductance can be used to implement centroid algorithm [3,4]. In Fig. 3 (a), M1-M15 constructs a conventional multiplier-type OTA. M16 and M17 generate the level-shift voltage to make M9 and M10 operate in the triode region. Under these conditions, the output current I_{OUT} of the multiplier-type OTA becomes:

$$I_{OUT} = 4KV_{O\_M1}(V_1 - V_O) = G_m(V_1 - V_O) \tag{1}$$

where K is the transconductance parameter of the M9 and M10. It can be seen that the transconductance of the OTA G_m is proportional to the control voltage V_{O_M1}. The output voltage V_O can be derived as [3]:

$$V_O = \frac{V_{O\_M1} \times V_1 + V_{O\_M2} \times V_2}{V_{O\_M1} + V_{O\_M2}} \tag{2}$$

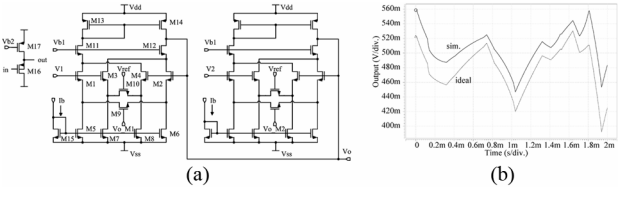


Fig. 3. COA defuzzification circuit by multiplier-type OTA

This circuit can be expanded for multiple inputs defuzzification circuit when the same block as M1-M15 is added. The simulation result of nine-input/one-output defuzzification circuit is shown in Fig. 3 (b). The mismatch between the threshold voltages of the NMOS and PMOS transistors introduces the output offset voltages. Removing it, the errors in the output voltages are less than 3% of the full-scale value.

3 Implementation of Neuro-fuzzy Chip

In this section, the previous functional blocks are combined into a two-input/one-output neuro-fuzzy system. Each input has three Gaussian-like membership functions. Six Gaussian circuits form Fig. 1, nine two-input/one-output minimization circuits

from Fig. 2 and a nine inputs defuzzification circuit from Fig. 3 are needed. Fig. 4 (a) and (b) illustrates the connection of these components and the microphotograph of the chip respectively. The supply voltage is ± 2.5V and the area is $0.78 \times 0.70 mm^2$.

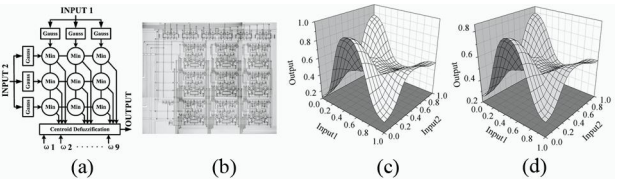


Fig. 4. (a) Two-input/one-output neuro-fuzzy, (b) microphotograph of the chip. (c) output surface by Matlab, (d) output surface by HSPICE.

Although there are six circuits from Fig. 1, only three different sets of Ve1 and Ve2 are needed for each input. They are (-0.2, 0.2), (0.3, 0.7), and (0.8, 1.2). The mean of Ve1 and Ve2 is the mean of Gaussian waveform, and the difference between Ve1 and Ve2 decides the breadth [3]. Their values are obtained by simulation to approximate the real Gaussian waveforms whose breadth is 0.25 and the means are 0, 0.5, and 1 separately. There are nine weights for this neuro-fuzzy system. Fig. 4 (d) shows the output surface of this neuro-fuzzy system compared with the Matlab simulation result (c). Here, the weights are 0.1V, 1V, 0.2V, 0.3V, 0.4V, 0.6V, 0.9V, 0.3V, and 0.5V. And the input ranges are designed from 0 to 1V. With a square wave input the output response shows a rise time of 300ns and a fall time of 280ns in 0.1V~0.85V. It corresponds to a speed of 3M FLIPS.

4 Conclusion

This paper proposes a low-power two-input/one-output analog CMOS neuro-fuzzy chip. Each analog blocks of this neuro-fuzzy system provides characteristics of high precision, high programmability, and compact structures. The neuro-fuzzy chip has been fabricated in 0.6-μm mixed-signal CMOS technology. It consumes a power of only 3.6mW with the complexity of 9 rules.

References

1. R. Navas, F. V. Verdú, and A. R. Vázquez: Neuro-fuzzy Chip to Handle Complex Tasks with Analog Performance. IEEE Trans. Neural Netw.. vol. 14 (2003) 1375-1392
2. H. Eichfeld, T. Kunemond, and M. Menke: A 12b General-purpose Fuzzy Logic Controller Chip. IEEE Trans. Fuzzy Syst.. vol. 4 (1996) 460-475
3. W. Z. Wang, D. M. Jin: CMOS Design of Analog Fuzzy System. J. Circuits Syst. Comput.. vol. 14 (2005) 1101-1112
4. S. Guo, L. Peters, and H. Surmann: Design and Application of an Analog Fuzzy Logic Controller. IEEE Trans. Fuzzy Syst.. vol. 5 (1996) 429-438

On-Chip Genetic Algorithm Optimized Pulse Based RBF Neural Network for Unsupervised Clustering Problem

Kay-Soon Low[1], Vinitha Krishnan[1], Hualiang Zhuang[1], and Wei-Yun Yau[2]

[1] Nanyang Technological University,
School of Electrical & Electronic Engineering,
Singapore 639798
k.s.low@ieee.org
[2] Institute of Infocom Reseach, Singapore
wyyau@i2r.a-star.edu.sg

Abstract. This paper presents a new technique for solving the unsupervised clustering problem using a pulse based RBF neural network that is optimized by a genetic algorithm. In this new approach, the neuron encodes the information using the firing times of pulses that are generated by the neurons. Due to hardware's speed advantage and its ability to parallelize along with reprogramability and reconfigurability, the field programmable gate array (FPGA) is used for implementing the proposed approach on-chip. The developed on-chip system is capable of solving various problems requiring large data set in real time. Experimental verification using a sample data set and Fisher's Iris data set has shown the feasibility of the developed system.

1 Introduction

GAs are adaptive heuristic search algorithm premised on the evolutionary ideas of natural selection and genetics. GA's have been applied to many optimization problems. Application of GA to increasingly complex problems could overwhelm software based GA and cause unacceptable delays. Thus, it may become a necessity to implement GA on hardware for solving complex real time problems. One of the key characteristics of GA is its repeatability, i.e. the search processes are repeated until optimal goals are reached. To ensure efficient hardware implementation of GA, repeatability is translated to reusability of hardware resources that results in speed advantage. Due to the inherent nature of parallelism in GA, FPGAs seems to be very apt for the GA application. Consequently, there is an increase in research in reconfigurable hardware systems and hardware GA [1]-[3].

GAs can be used to solve a wide range of problems. One such application is the unsupervised clustering problem. A clustering problem basically requires a method to partition a set of data points in such a way that similar elements or elements close to each other fall under one sub group and those dissimilar fall into others. Unsupervised clustering problem arise when we have a data set but they are not preclassified, i.e. no prior information about any natural subgroups they can form. Researchers [4]-[6] have shown that genetic algorithms can have notable success when devising set partitions.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 851 – 860, 2006.

In our previous work [7], a spiking neuron model that accepts input data, encodes and transmits the information with firing times of pulses has been developed. The design of the neuron dynamics makes the neuron yield a RBF relationship between its input vector and the triggering time of its output pulse with respect to the center (which is reflected in the synaptic delays). The proposed pulsed neuron approach can be used in various applications include clustering, function approximation among others. On account of parallelism in GA and its efficient hardware implementation, we use GA to optimize the synaptic delays to reflect the optimal cluster centers in this paper. When implemented on-chip, it could perform parallel processing of all data points and provides tremendous speed advantage.

The organization of the paper is as follows. Section 2 presents the basic concept of the spiking neuron and it's functioning. Section 3 deals with the application of GA to the unsupervised clustering problem. Section 4 focuses on the hardware implementation of the GA. In section 5 we discuss the results along with hardware usage. The algorithm has been verified with a sample data set of 68 inputs and the standard Fisher's Iris data set with 150 elements. Section 6 concludes the paper.

2 Spiking Neuron

The model of the spiking neural network used in this paper [7] is shown in Fig. 1. As shown in the figure, the output neuron, which functions as a radial basis function (RBF) unit, receives connections from the input neurons. The dashed segment between every input neuron to the corresponding synaptic terminal denotes the delay, which is defined as the difference between the firing time of pre-synaptic neuron and the time the post-synaptic terminal is activated.

A typical RBF is usually selected as a function of Euclidean distance between a center and the data point. To be consistent with this essential feature of the RBF, the spiking neural model is designed such that the synaptic delays d_{ji} will be selected so that their patterns reflect the center of input data points.

Consider that the input neurons receive an input data set **X** as

$$\mathbf{X}=\{X_k : X_k=[x_{k1},x_{k2},\ldots x_{kN}]^T,\ k=1,2,\ldots,n\} \tag{1}$$

where N is the dimension of the data and n is the size of the data set. Then the neurons encode the input information as firing times given by [7]

$$t_{kj} = \theta + T + \frac{1}{N}\sum_{i=1}^{N}\left(d_{ji} - c\cdot x_{ki}\right)^2 \tag{2}$$

where θ is a threshold value, T is a reference time, c is a constant, d_{ji} is the delay from the input to the post synaptic terminal. In this model, the output neuron will fire at the time t_{kj} when the integrated potential reaches a threshold θ. This equation yields an explicit RBF relation between t_{kj} and the Euclidean distance between the center

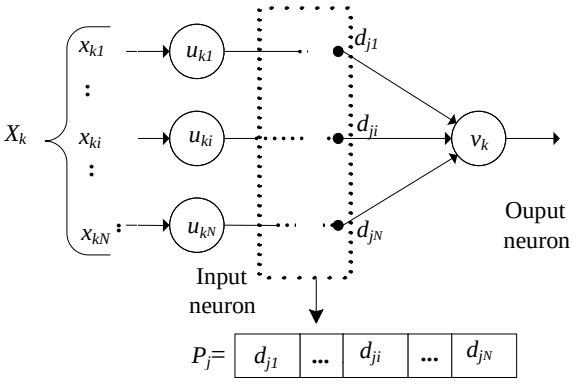


Fig. 1. Spiking neural network with single RBF unit

vector $[d_{j1},d_{j2},...,d_{jN}]^T$ and input vector $\left[x_{k1,}x_{k2,...,}x_{kN}\right]^T$. The shorter the Euclidean distance, the earlier the output neuron will fire.

3 Application of GA to Unsupervised Clustering Problem

3.1 Representation and Initialization

A genome (chromosome) is a set of genes. A gene is a set of possible alleles, and each of which codes for some characteristic. Consider the Fisher's Iris dataset [11] that has been widely used in cluster analysis as an example. This is a problem with four-dimensional inputs and three output clusters. The sepal length, sepal width, petal length, and petal width are measured in millimeters on fifty iris specimens from each of the three species, namely Iris setosa, Iris versicolor, and Iris virginica. Fig. 2 shows the chromosome representation of this problem. The chromosome is formulated by concatenating the integer codes of the features. As shown in the figure, each feature is represented by a synaptic delay d_{ji}. It uses 7 bits representation and yields an overall 28 bit chromosome. In general, a prior analysis of the feature set should be carried out to estimate the suitable number of bits required to represent the features to optimize the population size and memory in the chip design.

The GA population, whose chromosomes are concatenated strings of synaptic delays, can be expressed as

$$P=\left\{P_j:P_j=\left[d_{j1},d_{j2},...d_{jN}\right]^T,\ j=1,2,...,m\right\} \tag{3}$$

where m is the size of the population. The initial population of m chromosomes is generated using a random number generator (RNG) based on the multiple bit lagged Fibonacci number generator [10] implemented on the FPGA. It is also worth noting that the RNG is implemented such that it can be executed in parallel to all the other GA modules such as fitness evaluation, selection, crossover and mutation etc.

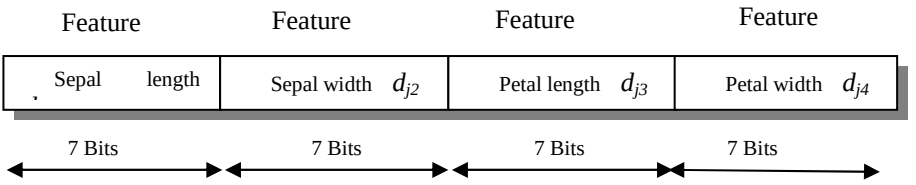


Fig. 2. Chromosome representation of Fisher's Iris data

3.2 Fitness Function Evaluation

Each input data is connected to an input neuron as shown in Fig. 3. All the input neurons are subjected to the same GA population. They are executed in parallel. The firing time of the output neuron is generated using (2). For each data input, *m* firing times are produced. Among them, the fastest time would successfully trigger the output neuron to fire a pulse. In this study, the fitness function of the chromosome is defined as the number of times that the chromosome is able to fire an output pulse for different data inputs.

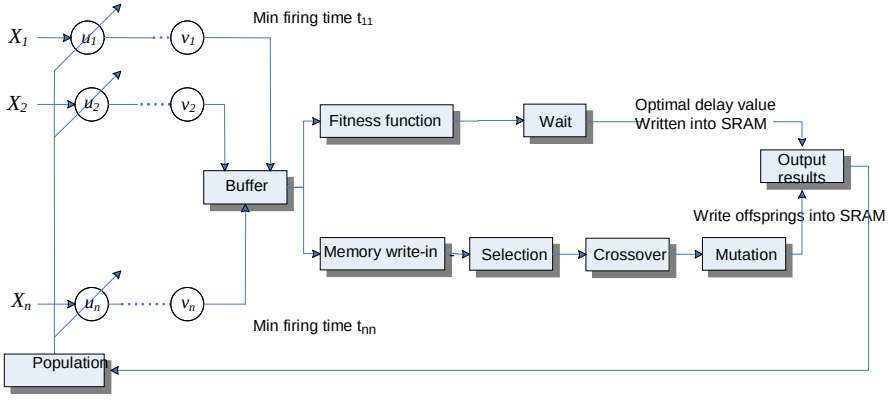


Fig. 3. Block diagram of GA based pulse neural network

Let α be a preset constant, which is functionally a firing time boundary. The fitness function evaluation for one generation can be summarized in Table 1. For each input data, there is a winning chromosome among the population. The fitness of the chromosomes are then evaluated and the highest rank chromosome, i.e. the winner of the generation, is determined. In simpler term, it implies that the delay associated with the winning chromosome is able to cluster a significant number of input data. After the fitness function evaluation, the optimal delay (P_{best}) and the associated inputs are stored for later use. To counter the possible effect of spurious clustering, the cluster radius or firing time boundary α is to be suitably selected. For unsupervised clustering, the near optimal α can only be determined after several training cycles with different α.

Table 1. Fitness function evaluation

%----------

Step 1: Start

Step 2: Initialize $C_{kj} = 0;\ k = 1,2,...n;\ j = 1,2,...m;$

Step 3: Parallel computations for all input data $k=1,...,n$;

Par

{

{ % First input data

j=1;

do

{ Determine t_{1j} using (2);

j=j+1;

} while(j<m);

If $\min_{j=1,2...m}\left(t_{1j} < \alpha\right)$, $C_{1j} = C_{1j} + 1;$

}

⋮

{ % n^{th} input data

}

}

Step 4: Determine fitness of each chromosome P_j

$$Fitness(P_j) = \sum_{k=1}^{n} C_{kj} \text{ , for } j=1,2,...m;$$

Step 5: Determine the best chromosome in the generation

$$\text{Compute } Max\left(Fitness\ P_j\right),\ P_{best} = P_j\ ;$$

Step 6: Store P_{best} and associated input values X_k which fired P_{best} .

Step 7: End

%----------

3.3 Selection and Mating

After the fitness of all the chromosomes are evaluated, the GA enters the selection and mating phase. In this phase, the probable candidates are first chosen for crossover to form the next generation offspring. The roulette wheel selection scheme for choosing the parents for mating has been used. The selection probability λ_j for each chromosome is given by

$$\lambda_j = \frac{fitness\left(P_j\right)}{\sum_{j=1}^{m} fitness\left(P_j\right)} \tag{4}$$

where the numerator $fitness(P_j)$ is the fitness value of the chromosome P_j and the denominator signifies the total fitness of the population. The number of copies that a particular chromosome is used in the selection pool is determined as

$$\gamma_j = \frac{fitness(P_j)}{\overline{fitness}} \tag{5}$$

where

$$\overline{fitness} = \frac{\sum_{j=1}^{m} fitness(p_j)}{m} \,. \tag{6}$$

Using (4) and (5), a new population of *m* chromosomes is derived and the system proceeds to the crossover stage. As shown in Fig. 3, the fitness evaluation and the selection stages are executed in a pipeline to improve processing time.

Fig. 4 depicts the crossover operation. As shown in the figure, the crossover of two individuals is performed on the individual feature independently, i.e. there are N independent crossovers per offspring. As all the parents are available after the selection phase, crossover operation can be performed in parallel on all the parents. This will result in all *m* offsprings being produced instead of 2 offspring at a time. The crossover probability is taken as 40%. After crossover, some of the chromosomes are randomly mutated by inverting some of their bit values. The mutation probability is kept very low at 4%.

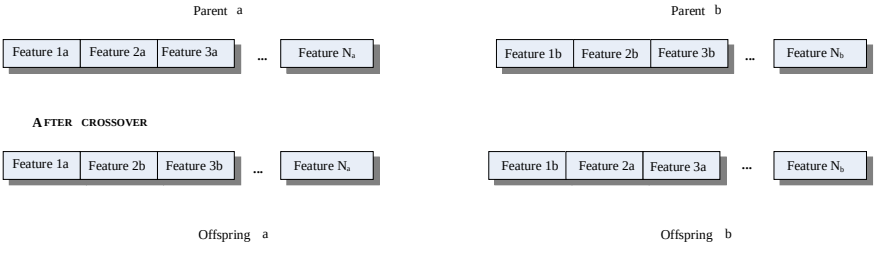


Fig. 4. Crossover operation between 2 chromosomes of dimension *N*

3.4 Association of Cluster Centers and Inputs

After the GA has undergone *K* generations, we have *K* cluster centers. To extract the cluster centers from the results, the following steps are followed:

1. For those cluster centers (and associated inputs) that have no similarity or are disjoint, they are first extracted.
2. Wherein two or more cluster centers group the same input data points, we discard all but one representative.
3. To deal with data points being associated with more than one cluster center, i.e. overlapping clusters, we calculate the Euclidean distance between the corresponding features in the data point and the delay elements they are associated with. For example, let the input data X_k be associated with *L* chromosomes. The *L* chromosomes are accumulated in a group G_L. The Euclidean distances between X_k and each of the chromosomes present in G_L is computed. The chromosome P_j that gives the minimum Euclidean distance is taken as the delay associated with the input data point X_k.

All the above steps are carried out until all the inputs are associated with a single delay element. The new population is then fed as new concatenated delays for the input neurons and the GA is repeated until all the inputs are associated with a center, or reaching 25 generations. Experimental results showed that 25 generations are sufficient for optimal clustering of the input data used in this study.

4 Hardware Implementation

Both the GA and the pulse neural network have been implemented using the Virtex II XC2V1000fg256 FPGA housed in a RC200 developer board. The FPGA has a maximum capacity of 1,000,000 logic gates and 5120 configurable logic blocks (CLBs). The RC200 board is fitted with two, 2-MB ZBT RAM banks.

GAs are in general highly parallelizable. In this paper, this feature is exploited to improve the system performance. There are a number of ways to parallelize GA [9]. In this paper, the master–slave model has been used. In the implementation, the master stores the population in memory and the slave carries out all the processes of the GA, namely fitness evaluation, selection and crossover/mutation operations. As the system can be implemented using a single FPGA, there is little communication overhead between the master and the slaves. Furthermore, both the on-chip and off-chip RAMs have been used for storing and retrieving the data parallely to yield better performance. A random number generator (RNG) based on the multiple bit lagged Fibonacci generator [10] is utilized to generate random numbers. The RNG is running in the background of the forward neural computations so that necessary random numbers are ready for the selection, crossover and mutation operations. During pipelining such as fitness function evaluation and selection stages, we need to ensure synchronism of the modules. Thus, it is necessary to have delay/wait modules to adjust the timings of the modules as in Fig. 3 so that the correct values are read in during every generation. The implementation is scalable, and we could accommodate very high population with the off-chip RAMs which has 2^19 address locations. It is estimated that a total of 1,000,000 chromosomes with 36 bits length could be accommodated using the proposed approach.

In the implementation, we have used the Handel C [12] as the coding language. The Handel C is a programming language that is based on ANSI-standard C. It is designed for compiling programs into hardware such as FPGAs or ASICs. Handel-C targets hardware directly, and provides several hardware optimizing features. At the basic level, the coding language's constructs can be exploited for parallel executing all simple statements. For example, it is used in our work for single value assignments and transferring the values from one array to another in just one clock cycle. The constructs can be written as

```
%.........parallel assignment for single value …….%
Par{count=0, j=1,k=1};
%.........parallel assignment for arrays ……%
Par(i=0;i<100;i++)
{ buffer[i]=fitness[i] ;
  buffer1[i]=delays[i];
}
```

Another use is the simultaneous calls to functions. Using arrays of functions, large blocks of functionality can be generated that execute in parallel. Based on this technical platform, parallel forward computations of all neurons and parallel genetic computations of all chromosomes are coded explicitly and compiled into digital hardware directly.

5 Results and Hardware Analysis

The developed system has been experimentally tested using a relatively small and simple set of data, namely a sample data set of 68 elements with 7 bits and the Fisher's Iris data. The key objective is to experimentally verify the functionality of the developed system before we expand it to full scale in future to other real time applications such as clustering of satellite images that have large data set.

5.1 Sample Data Set

In the proposed clustering algorithm, α is an equivalent measure of the predefined radius of clusters. Fig. 5a shows the sample data points for the testing and the resultant clustering with the proposed approach with α =10 units. The GA parameters are: population size *m*=100, number of generations=20, P_c = 0.4, P_m =0.04. As shown in the figure, the number of clusters after the GA optimization is 4. Fig. 5b shows the results with a different value of α =50 units, while all the other GA parameters being kept constant. At the end of GA optimization, the number of clusters is 2. These results are consistent with the previously published results [13].

5.2 Clustering of Fisher's Iris Data

The Fisher's Iris data set has been widely used for examples in cluster analysis. After running the experimental system, the developed system is able to cluster the data set in 25 generations. The clock speed of the experimental FPGA board has been set at 50

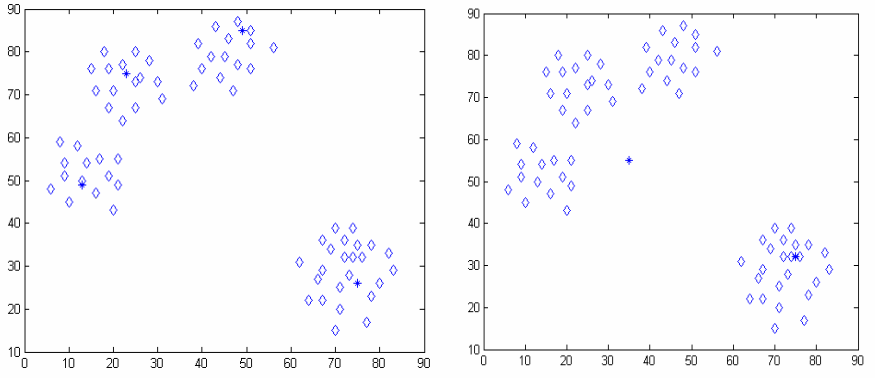


Fig. 5. (a) Clustering of sample data set with α =10 units **(b)** α =50 units

MHz and the timing analysis reveals that the minimum period of execution is 27.934 ns yielding a maximum frequency of 35.799 MHz. In general, the system takes approximately 2 msec to complete one GA generation. Overall, the system takes approximately 65 msec to complete the clustering of the Fisher's Iris data.

Table 2. Performance comparision

Method	Percentage of correct classification
Proposed method	90.99±.33%
k-means	88.67%±0.67%
SOM	87.33%±0.67%

Table 3. Hardware resouce ultilization

Logic utilization	
Slice flip-flops	1085(10% usage)
4-input LUTs	2206(21% usage)
Logic distribution	
Occupied slices	2037(39% usage)
Total number of 4-input LUTs	2400(23% usage)
Total equivalent gate count	220,928
Peak memory usage	107 MB

Table 2 compares the accuracy of the proposed approach with two other well known methods namely the k-means and SOM for clustering the Fisher's Iris dataset. The dataset can be divided into 3 clusters. Thus for the k-means method, k=3 and for the SOM approach, we use 3 Kohonen neurons. From the Table 2, the results show that the proposed approach yields a better performance of more than 2%.

Table 3 shows the logic utilization and distribution level. From the table, it is noticed that the usage of 4-input LUTs is only 23% and the gate count is 220,928 out of the 1 million available NAND gates.

6 Conclusions

This paper presented an on-chip implementation of a GA based pulse neural network scheme. The proposed approach has been applied to the unsupervised clustering problems namely, a sample dataset and the Fisher's Iris dataset. The results have shown that the proposed approach is feasible and has reasonably good performance. The developed system is very flexible and scalable. It can be extended easily for a larger data set (and/or multi-dimensional data) and its speed advantage makes it a potential practical approach for real time data clustering.

References

1. Aporntewan C., Chongstitvatana P.: A hardware implementation of the compact Genetic Algorithm, *Proc. 2001 IEEE Congress on Evolutionary Computation*, Seoul, South Korea (2001) 624-629
2. Hossam E. M., Ahmed Kadragi I., Yasser Hanafi Y.: Hardware implementation of genetic algorithm on FPGA, *Proc. NRSC 2004*, NTI (2004)
3. Wallace Tang and Leslie Yip: Hardware implementation of genetic algorithm on FPGA. *Proc. 47th IEEE Intl. Midwest Symposium on circuits and systems*, Hiroshima, Japan (2004) 549-552
4. Nasraoui O. and Krishnapuram R.: A Novel approach to unsupervised robust clustering using Genetic Niching. *Proc. 9th Int. Conf. Fuzzy Syst* San Antonio, TX (2000)
5. Falkenauer E.: Solving equal piles with the Grouping Genetic Algorithm. *Proc. 6th Int. Conf Genetic Algorithms*, San Francisco (1995) 492-497.
6. Greene W. A.: Genetic algorithms for partitioning sets. *Intl Journal on Artificial Intelligence Tools*, Vol. 10. (2001) 225-241
7. Zhuang H. L., Low K. S. and Yau W. Y.: A pulsed neural network with on-chip learning and its practical applications. *IEEE Trans. Industrial Electronics*, in press.
8. Hopfield J. J.: Pattern recognition computation using action potential timing for stimulus representations. *Nature*, Vol. 376. (1995) 33-36
9. Cantú-Paz E.: A survey of parallel genetic algorithms. *Calculateurs Paralléles, Reseaux et Systems Repartis.* Vol. 10, No. 2. (1998) 141–171
10. Chu P. and Jones R. E.: Design Techniques of FPGA Based Random Number Generator. *Proc. Military and Aerospace Applications of Programming Devices and Techniques Conf*. Laurel, MD. (1999)
11. Fisher R. A.: The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, Vol. 7. (1936) 179 –188
12. Handel C Language Reference Manual, Celoxica.
13. Zhuang H. L., Low K. S. and Yau W. Y.: A pulsed-based RBF neural model for unsupervised clustering. *Proc. Third International Conference on Computational Intelligence, Robotics and Autonomous Systems*, Singapore, (2005)

A Design on the Digital Audio Synthesis Filter by DALUT

Dae-Sung Ku[1], Phil-Jung Kim[2], Jung-Hyun Yun[2], and Jong-Bin Kim[2]

[1] Digital Broadcasting Examinations Division, Information and Communications Examinations Bureau, KIPO, 920 Dunsan-dong, Seo-gu, Daejeon, 302-701, Korea
dsku@kipo.go.kr
[2] Dept. of Electronics Engineering, Chosun Univ., 375, Seosuk-Dong, Dong-gu, Gwangju, 501-759, Korea
{yjhyun, jbkim}@Chosun.ac.kr, philjung@hanmail.net

Abstract. Semiconductor Technology and multimedia communication has been developed to interest in high quality video and multi channels audio. Especially, it is trend to important technology so that DVD market grows rapidly. Mpeg audio standard is available support for any bit rate. In this paper, the design of Field Programmable Gate Array(FPGA) is presented mpeg audio synthesis filter by Distributed Arithmetic Look Up Table(DALUT) method. In order to design for high speed filter that operated from 10MHz to 30MHz and DALUT is used instead of multiplier. In this paper, the output is designed by VHDL and is verification using C language. VHDL simulation used Model-sim SE 5.7f and synthesis is used Synplify Pro 7.2Ver. Target library is used XC4010E, XC4020EX, XC4052XL and P&R tool used XACT M1.4.

1 Introduction

The Mpeg audio is ISO/IEC standard method for High-quality and multi channels encoding. At resent times, all DVD and digital audio method are seeking multi channel and sound quality [1]. Unlike video compression offers 3 layers. Mpeg audio is used polyphase filter-bank, but it is supplementing fault by retreating output of filter using Modified Discrete Cosine Transform(MDCT). As different from polyphase filter-bank, MDCT part is able to remove aliasing that perform on reversibility conversion and offers better spectral resolution, and happen from polyphase filter part partially. And long block is 18samples to have different blocks and short block is composing of 6samples [2]. The filter-bank consists of a 50% overlapped critically sampled MDCT which can be switched between block lengths of 256 and 2048samples. The function of two blocks is as following. The length of long block has a better resolving power of frequency in audio signal which has a stable property, and the length of short block has a better resolving power of time domain in momentary signal [3]. The transform provided by compaction algorithm of MPEG audio has no loss in terms of perceptual hearing system of human beings. Additionally, it provides 6~12 times compression ratio [4].

The one of remarkable features of MPEG audio is the symbolization of subband which is the symbolization method to improve signal to noise ratio by dividing audio

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 861 – 869, 2006.

signal into several bands and lowering each sampling rate of each bands by decimation and then assigning a bit in proportion to the size of signal. Filter-bank could be split into polyphase and weighted overlap added filter-bank. MPEG audio is organized by the same number of sample in both time area and frequency area, and consecutive 32units of audio sample are processed as a one module [5][6].

However, in this paper the efficient hardware designed using DALUT method in the filter part of MPEG audio objectives. In PCM system the audio sampling frequency is formed as following table 1 and 32~224Kbps per a channel is generally used in bit rate.

Table 1. Sampling frequency of Digital Audio

Index	Frequency
00	44.1KHz
01	48KHz
10	32KHz
11	Reversed

In this paper, we designed MPEG audio filter part using DALUT method for hardware implementation. This paper is organized as follows. The theoretical background of DALUT is described in Sect. 2. Sect. 3 presents designed of synthesis filter and LUT for coefficient storage design. Experimental results and conclusions are given in Sect. 4 and 5, respectively.

2 Distributed Arithmetic Look-Up Tables

The arithmetic sum of products that defines the response of linear, time invariant networks could be expressed as equation. 1.

$$y(n) = \sum_{k=1}^{k} A_k X_k(n) \tag{1}$$

Where $y(n)$ indicates the response of network at time n, $X_k(n)$ indicates the Kth input variable at time n, and A_k indicates the weighting factor of Kth input variables. In filtering applications the constants, A_k , are the filter coefficients and the variables, X_k , are the prior samples of a single data source. In frequency transforming - whether the discrete fourier or the fast fourier transform- the constants are the sine/cosine basis functions and the variables are a block of samples from a single data source [7]. The multiply intensive nature of could be appreciated by observing that a single output response require the accumulation of K product terms [8]. We defining the number format of the variable to be 2's complement, fractional a standard practice for fixed point microprocessors in order to bound number growth under

multiplication. The constants may have a mixed integer and fractional format [9]. The variable, X_k, may be written in the fractional format as shown in equation 2.

$$X_k = -X_{k0} + \sum_{b=1}^{B-1} X_{kb} 2^{-b} \tag{2}$$

Where X_{kb} a binary variable and could be assume only values of 0 and 1. The time index has been dropped since it is not needed to continue the derivation. The final result is obtained by first substituting equation 1.

$$y = \sum_{k=1}^{K} A_k [-X_{k0} + \sum_{b=1}^{B-1} X_{kb} 2^{-b}] = \sum_{k=1}^{K} X_{k0} \bullet A_k + \sum_{k=1}^{K} \sum_{b=1}^{B-1} X_{kb} \bullet A_k 2^{-b} \tag{3}$$

Each term within the brackets denotes a binary AND operation involving a bit of the input variable and all the bits of the constants. The plus signs denote arithmetic sum operations. The exponential factors denote the scaled contributions of the bracketed pairs to the total sum. And the constructed Look-Up Table that can be addressed by the same scaled bit of all the input variables and can access the sum of the terms within each pair of brackets [10]. Fig. 1 shows that DALUT address and contents part.

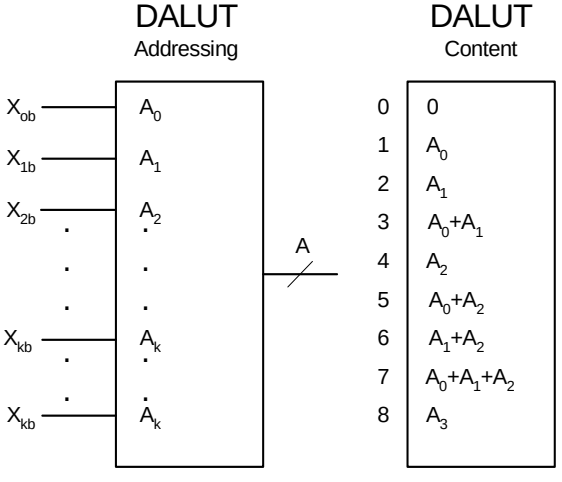


Fig. 1. The distributed arithmetic look-up table

The arithmetic operations have been reduced to addition, subtraction, and binary scaling. With scaling by negative power of 2, the actual implementation entails the shifting of binary coded data words toward the least significant bit and the use of sign extension bits to maintain the sign at its normal bit position. The hardware implementation of a binary full adder entails two operands, the addend and the augend to produce sum and carry output bits.

3 FPGA Implementation of Synthesis Filter

The synthesis filter requires the four arithmetical operations and accumulation, and in order to execute these operations it is represented by 2's complement and can run the operation. In a portion of MPEG audio decoder the part that has the most operation quantity is the synthesis filter. So in order to operate at high speed pipeline architecture is selected, and a multiplier is canceled by making the result of multiplication with cosine function into a table.

3.1 Design of Distributed Arithmetic Look-Up Table

In order to raise the rate, DALUT realized in the synthesis filter can reduce each table size by "Filter coefficient $\times 2^4$" by using shift addition of the result of multiplication between 4bit filter coefficient and hexadecimal 1digit. This composition is called parallel DALUT and the multiplication process of parallel distributed arithmetic LUT is described in fig. 2. Fig. 2 shows that the multiplier of parallel distributed arithmetic LUT.

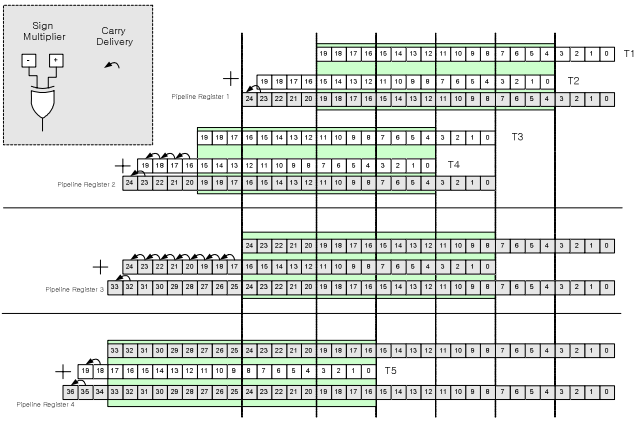


Fig. 2. Multiplier of parallel DALUT architecture

There is Look-Up table selector to make it possible for the selection of LUT which is suitable for block type of input signal and according to block type. The data array's size is 12 when the block type is 2, and in order case (0, 1, 3) there is the output address decoder which decodes a higher address by dealing with 36 data as a block and using 6bit counter. Allow 5 units of 4bits distribution input to make an output in accordance with a lower address. Five distributed arithmetic LUT output is generated by clock synchronization at the same time. It means that 21data is generated, and output data could be multiplied by 4bit shift addition.

All multiplication are generated as a positive real value so that the size of distributed arithmetic LUT is 1/2 and therefore the negative of input is calculated by using XOR according to a sign. Fig. 3 is the parallel pipeline multiplier including distributed arithmetic LUT, shift adder and accumulator. Whole blocks are operated by clock synchronization and have purpose of speed control, initializing, operation pin to

make the control of external controller convenient. Each LUT consists of one memory cell by using all the same memory space that table and table selector are combined into XC4010E. This block is for the solution of cosine function's multiplication. And the result of multiply synthesize window has must again operation and this also has the same architecture. It is processed within 24bit so that bigger bit is designed merely be able to calculate an accumulation result within 41bit.

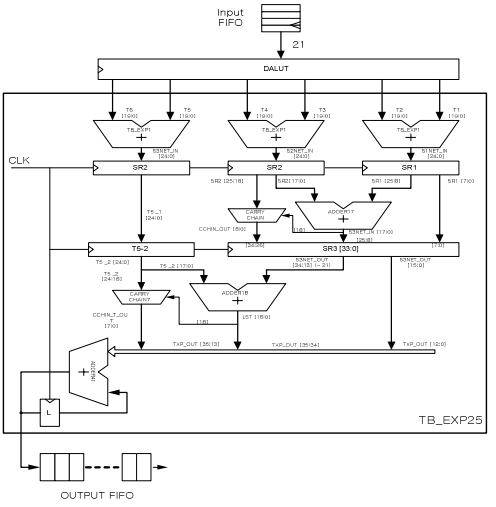


Fig. 3. Parallel pipeline multiplier using DALUT

Fig. 4 is representing LUT on the part of Digital audio decoder.

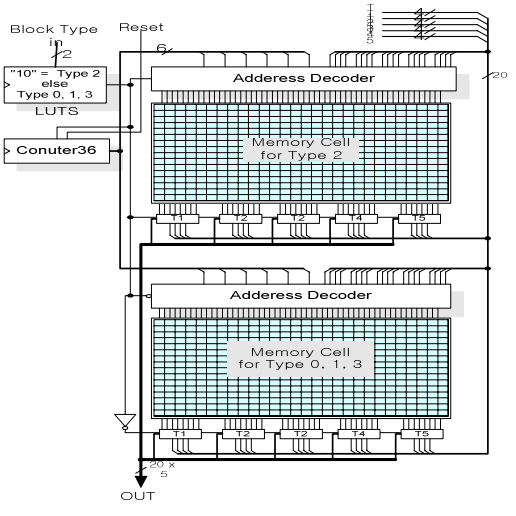


Fig. 4. Digital audio decoder LUT

3.2 LUT for Storage of Coefficients and Short Blocks Overlap Addition

In operating the parallel process LUT consists of one memory cell. And it is easy for LUT to be composed by multiple port memory only for reading. LUT can receive five addresses synchronously and specify respective five address space. Additionally it has five data buses so that it can generate five coefficient values synchronously. Because LUT receive a distribution signal sequentially, its can own control circuit within itself. Windows sequence is divide into Long(1024 samples) and Short(245 sample) that called a block type. In case of short block type, arrange 8 short windows into 1024 blocks because short windows overlapped addition. When short windows length exceeds 1024 of overlapped addition buffer because long windows and block length are different. Therefore, process dividing short windows 4th, fifth by nonlinear memory in criterion of short offset value. And others block to refill from first of overlap additions. Short offset value is toggled by 1 when is 0 and long state sequence in reset states. Fig. 5 shows that overlapped addition of short block.

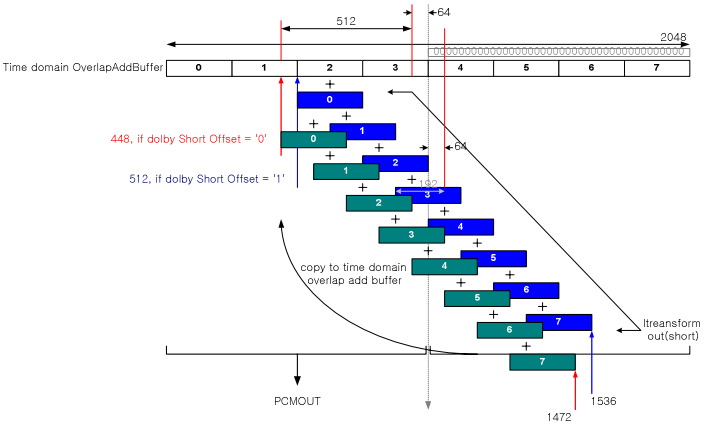


Fig. 5. Overlap addition of short block

4 DALUT Simulation and Verifications

4.1 Parallel Pipeline Design

LUT each output result could be added by using CSA. It is composed of 4bit RCA by basis group adder. And we designed so that can do carry select using multiplexer. In the part where a parallel process is impossible, it is designed to process the pipeline with a pipeline register located in the middle part. All the representation of numbers use the 2's complement and 1's complement could be converted into 2's complement if the adder receives a carry input. In the distributed arithmetic LUT, table and arithmetical operation part are synthesized respectively and the last character in arithmetical operation part represents a bit width. If whole part of LUT is realized by FPGA, it occupies a lot of CLB spatially so that a memory cell is stored in ROM and only

control block of LUT is synthesized FPGA. In case of simulation the value of ROM is verified by providing it with a component form. Fig. 5 shows that the parallel pipeline schematic architecture and describes the part which represents the whole hierarchy structure of synthesis filter using DALUT.

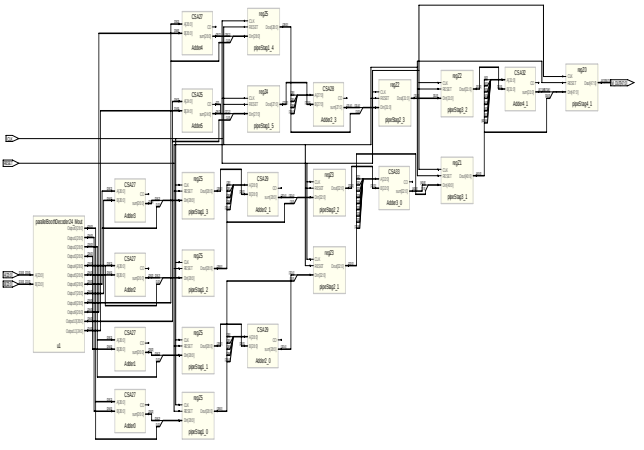


Fig. 5. Synthesis filter block diagram of parallel pipeline architecture

4.2 System Synthesis and Verifications

In this paper, the synthesize filter is the part which required to be realized. It could be realized by port mapping all blocks designed in sub blocks and connections. Fig. 6 describes the part which represents the whole hierarchy structure of synthesis filter using DALUT. Fig. 6 represents simulation result of the entire system.

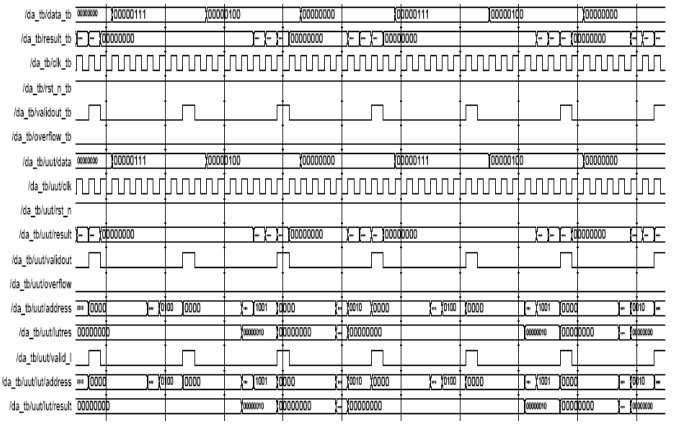


Fig. 6. Simulation result of synthesis filter

In this paper, used XILINX FPGA resource, table2 represents FPGA sources which compose entire synthesis filter.

Table 2. XILINX FPGA design resources of synthesis filter

FG Function Generators	1274
H Function Generators	253
Number of CLB cells	973
Number of Hard Macros and Other Cells	274
Number of CLBS Other Cells	322
Total Number of CLBS	1295
Number of Ports	93
Number of clock pads	0
Number of IOBs	0
Number of Flip Flops	1496
Number of 3-state Buffers	0
Total number of Cells	1247

5 Conclusions

In this paper, the synthesis filter part is realized by FPGA using DALUT method. To realize a high speed filter a DALUT method is used instead of a adder so that the filter is able to operates at least 30MHz and the maximum frequency of the cheap designed operates on 70MHz. There is 30% improved in aspect of the operation speed. Because DALUT structure doesn't need a multiplier, a memory that has a large capacity is required instead of a high speed processor. All system modules are designed by VHDL so that it is very easy to apply for ASIC. The realized system is divided by several FPGA and synthesized but when using ASIC, synthesize efficiency is improved so much that the size becomes much smaller significantly. And due to improvement of speed, actually it can decode more than 50channels. The high level computers capacity and long time are required for entire simulation and the verification of result is also very difficult works.

The verification is operated by comparing software composed by C language with simulation result that is using ECAD tools. It stores the file of 2's complement output by using text-IO of VHDL. In this paper, the suggested synthesis filter of DALUT method could be easily applied to DSP chip used in sound codec of digital audio.

References

1. Davis Pan "A Tutorial on MPEG Audio compression" Motorola Inc.
2. KEN C. POHLMAN, "Principle of Digital Audio 3rd Edition", McGraw-Hill 354-415, 1997.
3. J. Princen, A. Johnson, A. Bradely, "Subband/Transform Coding using Filter bank designs based on Time Domain Aliasing Cancellation", Proc, of the ICASSP 1987, pp.2161-2164.

4. The sorer, Kh. Brandenburg, “The use of multirate filter banks for coding of highquality digital audio” 6th European Signal Processing Conference, Amsterdam, Vol.1, pages 211-214, June 1992.
5. Ted painter, Andreas Spanias, “A Review of Algorithm for Perceptual Coding of Digital Audio Signals” Department of Electrical Engineering, Telecommunications Research Center Arizona State University.
6. P.P. VAIDYANATHAN, “Multirate Systems and Filter Banks”, Prentice Hall Englewood Clieffs, 1993.
7. ALEXANDER D. POULARIKAS, “The Transforms and Applications Hand Book” CRC and IEEE PRESS, 1996.
8. ISO/IEC JTC1/SC29/WG11 MPEG, International Standard IS 13818-3 “Information Technology Generic Coding of Moving Pictures and Associated Audio, Part 3: Audio”, 1997.
9. Vijay K. Madisetti “VLSI digital signal processors an Introduction to Rapid Prototyping and Design Synthesis”, IEEE Press.

Video Encoder Optimization Implementation on Embedded Platform*

Meng Qinglei, Yao Chunlian, and Li Bo

Digital Media Laboratory, Beihang University, China
mql_198029@tom.com

Abstract. To satisfy the requirement of embedded video encoder in various scenarios, an embedded universal video coding system based on embedded processors is proposed in this paper, which combines the advantages of multi processors in single architecture. According to different channels, two feasible solutions are given to avoid the overflow and underflow of buffer while transmitting the compressed data. Due to limited computational resources and real-time application requirement, each module of the MPEG-4 codec is reviewed and efficiently implemented based on the embedded hardware platform. Experimental results show that the embedded system can successfully be used for real-time application.

1 Introduction

With the rapid development of Internet and mobile communication, digital image communication has played a more and more important role in multimedia services. But one problem occurs according to the digital video, it's mass of data. So reducing the transmission bit-rate and concomitantly retaining video quality has become a challenge for efficient video compression. As we know, there are several established international standards, such as H.263 [1], MPEG-4 [2] and JVT [3]. As one of the best compression standards, MPEG-4 is adopted frequently for many applications. In recent years, with the advent of high speed and low power embedded processors, the increasing demands for embedded real-time digital video application, which mainly based on programmable devices, like Digital Signal Processor (DSP), RISC, FPGA, ARM et al, have attracted tremendous attention. Embedded processors process data stream by specific algorithms that take advantage of the high integration scale, robustness and flexibility of the digital circuits. However, due to the limitation of chip resource and complexity of algorithm, efficient algorithm and implementation is necessary to achieve real time.

The rest of this paper is organized as follows: In Section 2, we give a systematic view of an embedded video system and point out some key points when practicing a similar embedded system. Section 3 explains the main idea of MPEG-4 video

* This work is partially supported by the Specialized Research Fund for the Doctoral Program of Higher Education, the National Defence Basic Research Foundation, and the Program for New Century Excellent Talents in University, and the research was made in the State Key Lab of Software Development Environment.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 870 – 879, 2006.

compression techniques. In Section 4, some implementation and optimization details are described, and experimental result based on a test platform is provided. Finally, concluding remarks are given in Section 5.

2 System Description

Generally speaking, embedded processors have such characteristics as high-speed performance frequency, high-speed memory bus, flexible memory controller and the ability to handle variable-latency I/O devices. They sustain multi-operand parallel instruction set and special DMA bus. With the higher integration, some embedded processors have transmission port, such as Multi-channel Buffered Serial Port (McBSP) and Ethernet-network MAC, in the chip.

Fig.1 shows the whole data flow of the universal system made up of three modules: encoder, decoder and transmission/ receiving. On the encoding side, raw live video data is captured and then stored into an input buffer in a given format like YUV 4:2:0. The encoder receives extern instruction to do initialization and adjustment first, then encodes the video data and stores the compressed bit stream into an output buffer. The transmitter takes the task of getting data from output buffer and then puts all these bits to channel with a constant/variable speed. On the decoding side, the receiver reads data from channel and writes into the receive buffer. The decoder gets data from the receiving buffer if available and decodes them. Finally, after the post-processing, decoded video is displayed on the screen. We can see that the whole architecture of encoder and decoder is symmetric. Embedded processors are the kernel of the framework. So we can design two similar hardware platforms except the transmitter and receiver. There are two kinds of channel type, which are Variable Bit Rate (VBR) and Constant Bit Rate (CBR), so we proposed two feasible solutions to avoid the overflow and the underflow of output buffer while transmitting the compressed data.

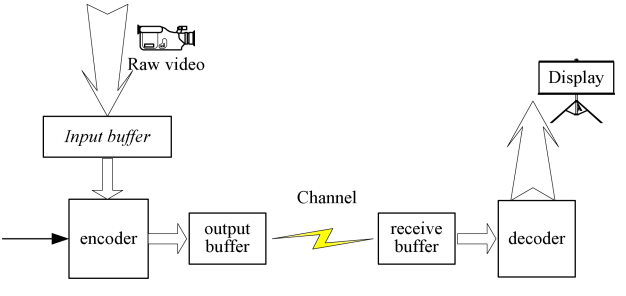


Fig. 1. System architecture

1. VBR: variable bit rate like wireless LAN, GPRS

We design a circular buffer as output buffer as is shown in Fig.2 (A). The buffer is divided into several parts. Each compressed frame of video sequence is written into each part.

2. CBR: constant bit rate like McBSP

In the system, the compressed bits are written into the output buffer at an inconstant speed, while the system sends the data to channel at a constant speed. So, according to the difference between the two speed, we design and accomplish a software FIFO (First in First out) to avoid the output buffer overflow and underflow, which is shown in Fig.2 (B).

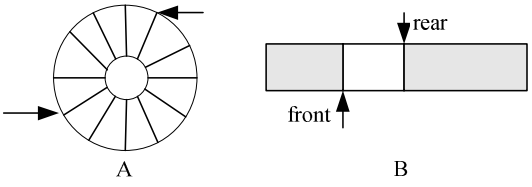


Fig. 2. Output buffer diagram

- Avoid buffer overflow

Before compressing the next frame, we check the empty space of buffer that whether the compressed bits can be stored or not. If not, encoder will not compress the next frame and wait for a short time until there is enough space.

- Avoid buffer underflow

After writing the compressed bits, we check the fullness of the buffer whether it has achieved a possible underflow threshold or not (Generally, the underflow threshold value is set to 10%). If so, a certain amount of "0" will be filled into the output buffer until 10% space is captured.

3 MPEG-4 Review

MPEG-4 standard aims at providing standardized core technologies allowing efficient storage, transmission and manipulation of video data in multimedia environments. It employs a Motion Compensation based DCT coding scheme. As is shown in Fig.3, it consists of DCT, ME/MC, Quantization, VLC, Rate Control and Image reconstructed. The coded frame type can be intra frame (I), forward predict frame (P) and bidirectional prediction frame (B), and P, B frame can also be called inter frame. The basic encode unit is 16×16 Macroblock, and then Macroblock is divided into 8×8 blocks. The details are explained as follows:

DCT (Discrete Cosine Transformation): DCT is a reversible transformation used to map spatial image into a set of frequency coefficients. It is well known that different DCT coefficients represent different information of image. Usually, low frequency coefficients can tell the silhouette of image and high frequency coefficients can tell the detail information of image, which has small magnitudes and can be coarsely quantized or discarded with little image distortion.

Quantization: According to human vision system, the quantization matrix is used, and each DCT coefficient is divided by the value on the corresponding position on the matrix to reduce or eliminate some of unimportant data.

VLC (Variable Length Code): The 8x8 blocks of transform coefficients are scanned with "zigzag" scanning. A three dimensional variable length code (LAST, RUN,

LEVEL) is produced by run length code. With a fixed-probability code table, Huffman coding is used to code coefficients.

RC (Rate Control): Rate control in video coding has a significant influence on the bits and image quality. The goal of rate control algorithm is to distribute the coded bits properly to each Macroblock at a given total bits budget. Generally algorithm adopts a Rate-Distortion model to calculate the quantization step.

Image reconstructed: After compressing one frame, the coefficients are decoded and copied to the reconstructed frame buffer as a reference frame for prediction. This process consists of IDCT, filter, frame buffer and Inverse quantization.

ME/MC (Motion Estimation and Compensation): For inter frame, the current Macroblock/block finds out the best match prediction block with the minimum SAD value in search window of prediction frame. And the distance between current block and prediction block is called motion vector. The core of algorithm is search strategy and matching criteria to get the motion vector.

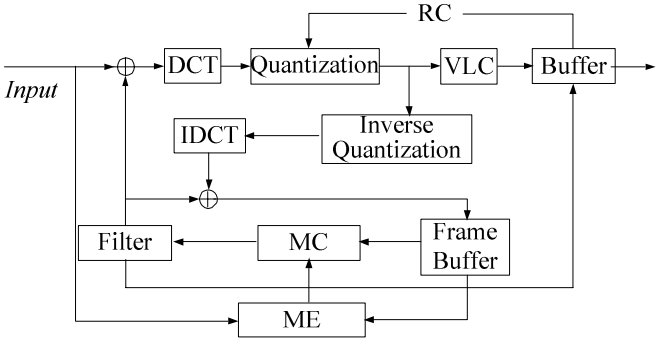


Fig. 3. The structure of the mpeg-4 video encoder

4 Software Optimized Implementation

Our implementation is on such embedded hardware platform, shown in Fig.4, which is based on TMS320DM642 of TI cop and FPGA. TMS320DM642 is a fixed-point DSP with 600MHz frequency, 256KB RAM and the CPU consists of 8 parallel processing units. Some system level optimization methods are used, which are all described in my previous paper [6]. Experimental results show that these methods enhance the performance of the encoder more than 30 times. Corresponding to each function in section 3, some other major optimized implementation issues are discussed:

4.1 DCT/IDCT

TI Corp provides IMGLIB (image/video processing library) and DSPLIB [7],two sets of assembly-optimized key functions for data processing. By reasonably utilizing available resources and avoiding potential resource conflicts, each function in the two libraries is designed to achieve the best performance, such as IMG_fdct_8x8 and IMG_idct_8x8.

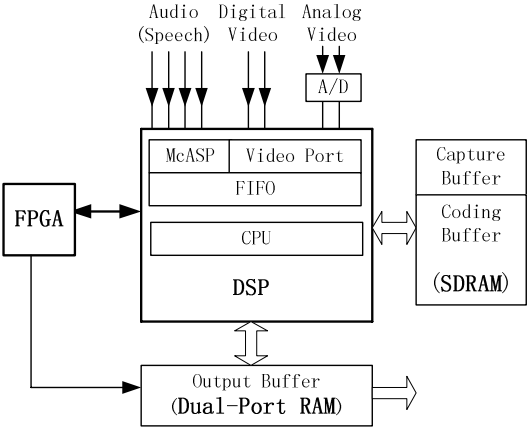


Fig. 4. Hardware Framework

4.2 Quantization/ Inverse Quantization

Quantization is the sole part that introduces distortion in compression. In MPEG-4 standard, a two-step quantization is adopted. Since human visual characteristics are considered, the two-step quantization obtains better subjective quality. But because of the division and floating-point computation, the encoding complexity is increased at the same time (especially in the embedded systems based on fixing-point processor).

To solve these problems, an efficient quantization computation method (EQ4DSP) based on fixed-point DSP has been proposed. The reference [8] describes the details. Here is the example in inter mode:

The quantization in MPEG can be described as Eq. (1).

$$LEVEL = \frac{16}{2 \cdot QP \cdot Quant\_Mat} \times COF \tag{1}$$

We transform Eq.(1) as:

$$LEVEL = \left(\frac{16 \cdot 2^N}{2 \cdot QP \cdot Quant\_Mat} \times COF \right) >> N \tag{2}$$

Finally, we obtain:

$$LEVEL = (Quant\_Self \times COF) >> N \tag{3}$$

$$Quant\_Self = \frac{16 \cdot 2^N}{2 \cdot QP \cdot Quant\_Mat}$$

Where "$>>$"denotes right shift, N denotes bits of right shift. $Quant\_Self$ is new pre-constructing shift quantization matrix for every QP.

By equivalently transforming original quantization equations and pre-constructing shift quantization matrix, complex division and floating-point computations are

replaced by simple fixed-point multiplication and shift operations. Furthermore, the conditional criterion and branch jump in the saturation can be avoided. So the efficiency of quantization is improved greatly.

But it still can not meet the real-time coding requirement, so we optimize the procedure at C-code level. Main optimization methods based on our experience consist of the following parts:

1. The first optimization step that can be performed on the DSP platform is to use compiler options. To get better performance, we need trade off the improvement of execution speed and incresement of code size.
2. Using intrinsic function to replace complicated C/C++ code and the packed data processing to maximize data throughput. The C6000 compilers support special functions called intrinsic operators, such as _add2, _mpy, _lddw, _mem4_const, etc. those intrinsic operators can be translated into efficient assembly code and can improve the execution speed,
3. Unrolling loop to increase software pipelines. Since the compiler only evaluates the possible loop optimization at the top of loops, we need expand small loops manually and unroll inner loop.
4. Replace the multiply and division by the logic shift because of higher speed.

To improve efficiency further, we use linear assembly to rewrite the key functions. The compiler allows writing linear assembly code without being concerned with the pipeline structure or registers assignment. The compiler assigns registers and uses loop optimization in order to turn linear assembly into highly parallel assembly.

Take the quant_intra function for example, after optimization the core loop code has higher parallelization that even achieves seven instructions parallel in one clock cycle. The result is shown in Table 1 and more details are described in subsection 4.6.

4.3 VLC

The 8x8 blocks of transform coefficients are scanned with “zigzag” scanning. A three dimensional variable length code (LAST, RUN, LEVEL) is produced by run length code. With a fixed-probability code table, Huffman coding is used to code coefficients.

The code table composed of (Run,Level,Last) is an typical unsymmetrical binary tree, and is sparse at the root, so the whole table will consume lots of memory. Due to the limitation of the memory resource in our system, we divide the whole table into several small tables. The index item of each code table represents table’s memory space. We get the scanning result and use look-up table to find the item position in table, and then get the code. These small tables can reduce memory space and be suitable for embedded application.

4.4 Rate Control

To achieve the requirements of low delay and high efficiency, the low delay rate control algorithm is proposed in my paper [9]. It uses slice bits allocation method based on histogram of difference image and Rate-Distortion control strategy.

The HOD between frame m and n is provided as follows:

$$HOD(f_n, f_m) = \frac{\sum_{i>|TH|} hod(i)}{N_{pix}} \tag{4}$$

Where i is the index of quantization bin, Here hod(i) is the histogram of the difference image between fn and fm, Npix is the number of pixel, TH is the threshold value for detecting the closeness of the position to zero.

With the HOD value, the target bits of the slice T_{kslice} can be calculated with formula (5).

$$T_{kslice} = \left(1 + \mu \times \frac{HOD_k - \frac{1}{k}\sum_{i=1}^{k} HOD_i}{\frac{1}{k}\sum_{i=1}^{k} HOD_i}\right) \times \frac{\left(\frac{B_{rem}}{N_{rem}} - R_{used}\right)}{N_{remslice}} \tag{5}$$

Here T_{kslice} is the target bits for slice k, HOD_k is the HOD for slice k in current frame, B_{rem} is the remaining bits before encoding current frame, N_{rem} denotes the remaining frame number. $N_{remslice}$ denotes the number of remaining slice in current frame and R_{used} is the used bits in current frame.

The quantization step QP for the current slice is calculated with the R-D formula (6).

$$T = \mathrm{X2} \times MAD \times QP^{-2} + \mathrm{X1} \times MAD \times QP^{-1} \tag{6}$$

The encoding bits are denoted as T. The image complexity is mean absolute difference (MAD) of overall slice. The quantization step is denoted as QP. The model parameters are denoted as X1 and X2.

Experimental result demonstrates that the improvement of architecture and algorithm can enhance the subjective image quality and decrease the overall delay.

4.5 Image Reconstruction

In the system, image reconstruction consists of IDCT, filter and frame buffer and Inverse quantization. The improved parts are quite similar to DCT and quantization.

4.6 ME/MC

ME algorithm consists of two stages, integer-pixel accuracy motion estimation (IME) and half-pixel accuracy motion estimation (HME).

To improve efficiency and avoid above defects, we put forward the improved algorithms which are proposed by our research group: smooth motion vector field adaptive search technique (SMVFAST) [12], and fast sub-pixel motion estimation algorithm* (FSME*) [13].

Furthermore, FSME* directly calculates the best position according to the neighboring integer sample value and the intermediate result of IME. The algorithm is mainly based on that a 2-D motion can be divided into two 1-D (horizontal and vertical direction) motions, and then the best position can be calculated respectively at the horizontal and vertical direction. FSME* experimental results show that it obtains similar PSNR to HME with only 0.02dB-0.07dB quality decrease on average and nearly 50% time of HME drop, and what's more FSME* don't need additional memory space for half-pixel image.

After implementing algorithm, we optimize the system further at code level. For example: sad16 function with two loops used to calculate the SAD value, interpolate function (InterpolateH, InterpolateV, InterpolateHV) used to perform the reference region interpolation and transfer function (Transfer16to8sub, transfer_16to8copy, etc) used to achieve the compensation. These functions spend mostly time of ME/MC processing. So we optimize these by using linear assembly.

The following is a linear assembly sample of sad16 function partly with two loops. We unroll inner loop and use the extended packed data intrinsic operation like _ldndw, _stdw and the intrinsic operators like _subabs4, _dotpu4. The optimization result is given in Table 1.

Original C code:

```
for (j = 0; j < 16; j++) {
     for (i = 0; i < 16; i++) {
          sad += abs(*(ptr_cur + i)- *(ptr_ref + i));
      }
          ptr_cur += stride;
          ptr_ref += stride;
}
```

Linear Assembly code:

```
LDDW            * A_srcImg ,     A_s7654:A_s3210
LDNDW           * B_refImg ,     A_r7654:A_r3210
SUBABS4         A_s7654,         A_r7654,    A_d7654
SUBABS4         A_s3210,         A_r3210,    A_d3210
DOTPU4          A_d7654,         A_k1,       A_s1
DOTPU4          A_d3210,         A_k1,       A_s0
ADD             A_sad,           A_s1,       A_sad
ADD             A_sad,           A_s0,       A_sad
```

We optimize original C code by unrolling inner loop and utilizing assembly instruction. In Linear assembly code, LDDW and LDNDW instruction can load 64-bits data from memory and store into register at 5 cycles.SUBABS4 can Calculate the absolute value of the differences for each pair of packed 8-bit values.DOTPU4 can multiply each pair of 8-bit values in operand and sum four products.

In order to refine linear assembly code, we consider the function unit and register assignment. Fig 5 shows the function unit and registers assignment. DSP has two side ways register, A-side and B-side, Each side has 16 registers. We assign A-side register for LDDW instruction and B-side register for LDNDW instruction. LDDW .D1 uses the data path D1,A-side register, LDNDW .D2 uses the data path D2, B-side

register. There is no conflict of register and function unit, and the performance of the sad16 function is improved greatly.

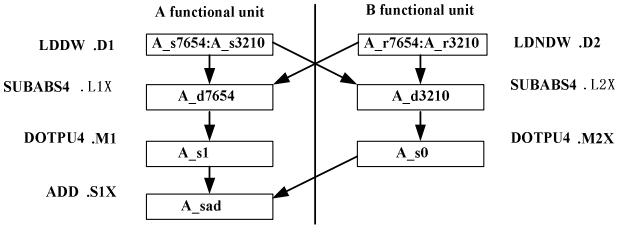


Fig. 5. Functional unit and registers assignment

Table 1. Comparison of run cycles on DSP

Function	Fdct	InterpolateH	Sad16	Transfer 16to8sub	quant_intra
Original	9342	16488	8760	5523	5126
C-Code	---	1757	1851	1835	---
Assembly	220	131	132	129	166
Futher	---	125	121	119	---
Multiple	42.46	131.90	72.39	46.41	30.88

Table 1 lists some results of code optimizations, and fdct use the library of IMGLIB. Experimental results show that code optimizations improve the performance notably; the executive cycles of some functions improve more than 100 multiples.

After optimization and implementation, the final result [6] of the project in Table 2 can achieve real-time application for the encoding with the multi-channel CIF format.

Table 2. Performance result

	Non-Optimization	Optimized
Frame Rate	0.32~0.37	25.58~27.61

5 Summary

In this paper, we describe an embedded multimedia coding hardware platform based on embedded processors, which always combine multi processors in a single architecture and consist of encode, decoder and transmission/receiving modules. Considering the channel with a constant/variable speed, two feasible solutions that output buffer design and buffer control strategy are given to avoid the overflow and the underflow while transmitting the compressed data. And then by in-depth studying MPEG-4 standard and its verified model, each modules of codec are efficiently implemented based on an actual hardware platform. Experimental results prove that the actual system can encode video efficiently and achieve multi-channel real-time application.

References

[1] ITU-T Draft Recommendation H.263, Video coding for low bitrate communications. Dec. 1995.

[2] MPEG-4 Video Verification Model version 15.0 [S]. ISO/IEC JTC1/SC29/WG11 N3093, Maui, Dec 1999.

[3] T. Wiegand, G. Sullivan. Working Draft Number 2, Revision 8 (WD-2, rev 8). Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, April 2002.

[4] Zhao Baojun, Shi Caicheng, Bi Li et al. Implementation of Real Time 2D DCT with FPGA and DSP, Acta Electronica Sinica .2003Vol.31,No.9,sep.2003:1317-1319.

[5] Jicheng Chen, Qingdong Yao, Peng Liu, Ce Shi, DSP with Some RISC Feature for Embedded System, ICSP'2004 Proceedings

[6] Wei Li, Qing-Lei Meng, Chao-Gang Wu, An Embedded 4-Channel Real-Time Multimedia Coding System, International Conference on Multimedia Modelling, MMM2006, January 2006

[7] Spru190.pdf,TMS320C6000 Peripherals Reference Guide

[8] Li Wei. Research of video coding and Dsp-based implementation [Ph.D. dissertation]. Beihang University, Beijing 2003

[9] Qinglei Meng, Chunlian Yao, and Weijin Huang, Low Delay Rate Control Algorithm for an Embedded Video Encoder,1st International Symposium on Systems and Control in Aerospace and Astronautics, ISSCAA 2006, January 2006

[10] K. K. Ma, P. I. Hosur, "Performance Report of Motion Vector Field Adaptive Search Technique (MVFAST)". In: ISO/IEC JTC1/SC29/WG11 MPEG99/m5851, Noordwijkerhout, NL, Mar. 2000.

[11] A.M. Tourapis, O. C. Au, and M. L.Liou, "Fast Block-Matching Motion Estimation using Predictive Motion Vector Field Adaptive Search Technique (PMVFAST)", In: ISO/IEC JTC1/SC29/WG11 MPEG2000/m5866, Noordwijkerhout, NL, Mar. 2000.

[12] Li Wei, Zhou B, Li B, A Fast Motion Estimation Algorithm Using Adaptive Motion Vector Field Search, Chinese Journal of Computers, 2003, 26(2): 168-173.

[13] Zhang Jinyin. Block Matching Sub-pixel Motion Estimation and Its Application [M.S.dissertation]. Beihang University, Beijing 2004

Effect of Steady and Relaxation Oscillation Using Controlled Chaotic Instabilities in Brillouin Fibers Based Neural Network

Yong-Kab Kim[1], Soonja Lim[1], and Dong-Hyun Kim[2]

[1]School of Electrical Electronics & Information Engineering, Wonkwang University, 344-2. Sinyong-Dong, Iksan, Chon-Buk 570-749, Korea
ykim@wonkwang.ac.kr
http://www.forlab.wonkwang.ac.kr
[2]Department of Mechanical Engineering, Wonkwang University
344-2 Sinyongdong, Iksan, 570-749, Korea
ncatkim@hanmail.net

Abstract. In this paper the effect of steady and relaxation oscillation in optical fiber based on neural networks and hardware implementation is described. The inherent optical feedback by the backscattered Stokes wave in optical fiber leads to instabilities in the form of optical chaos. The controlling of chaos induced transient instability in Brillouin-active fiber based smart structures leads to neural networks with steady and relaxation oscillations. Controlled chaotic instabilities can be used for optical logic and computations and optical sensing. It is theoretically possible to apply the multi-stability regimes as an optical memory device for encoding and decoding and complex data transmission in optical systems. These memories can be estimated as an optical logic function used for all-optic in-line switching, channel selection, oscillation, optical logic elements in optical computation with neural network application.

1 Introduction

Optical fibers based on neural networks application and hardware implementation have been extensively used in optical systems [1], [2]. Recent interest has been also focused on using optical fibers as sensors since fiber parameters are sensitive to the fiber immediate environment [3]. Specially, in the case of stimulated Brillouin scattering (sBs) sensor, the backward scattering nature of scattering has long been viewed as an ultimate intrinsic loss mechanism in long haul fibers since Brillouin threshold decreases with increasing effective fiber length. On the other hand, the very backscattering nature of this process and the existence of a threshold, provide potential optical device functions, such as optical switching, channel selection, amplification, sensing, arithmetic and neural functions in optical signal processing, and neural network applications and hardware implementation.

Active device in optical systems generally require the employment of nonlinearity, and possibly feedback for increased device efficiency. The presence of nonlinearity together with intrinsic delayed feedback has been repeatedly demonstrated to lead to

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 880 – 883, 2006.

instabilities and optical chaos [4], [5]. This phenomenon has extensively investigated by us for its potential detrimental effect to the Brillouin fiber sensor [6], [7].

An artificial neuron, used in neural network research, can be thought of as a device with multiple inputs and single or multiple outputs in hardware implementations. The inputs to a neuron are weighted signals. Neuron-type operations can be performed by an optoelectronic system that uses sBs for the weighted summation required in a neuron. Weighting can be achieved by optical summation and subtraction, conveniently carried out in an optical fiber using sBs. Weighted additions and subtractions are needed in many situations. For example, a neuron performs weighted summation of the incoming signals. The performance of such a device will enhance if it operates optically. We propose to study a system that can perform the practical implementation of a Brillouin-active fiber for optical neural net, neural function by exploiting the acousto-optic nature of the sBs process [6], [7].

2 SBS Network Implementation

A practical sBs logic implementation of theoretical neuron based neural networks and hardware implementations, as shown in Figure 1.

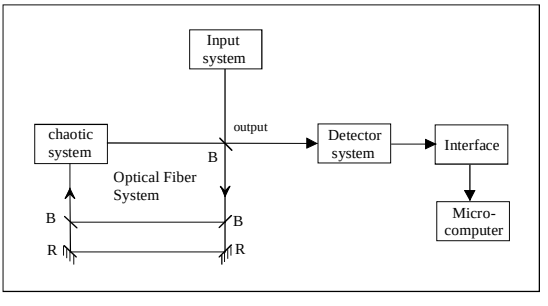


Fig. 1. Schematic diagram for controlling chaos induced instability in optical fiber based neural network system. The optical implementation included a chaotic system configured.

A stabilized *cw* probe laser operating was used as a pump source for low scattering losses in the fiber. Some levels of temporal instability and chaotic behavior in the backscattered intensity and also in its spectral line shift have been observed(see Fig.2). It is thus essential to know whether insertion of an amplifier will further destabilize the optical system. When the pump power reaches a threshold value, a temporal structure arises in the backward signal, consisting of a periodic train of Brillouin-wave pulses as shown in Figure 2(a). The Brillouin pulse train amplitudes remain unstable, particularly just below pump threshold. When the observation is made using a long time scale (100μsec/division), the Brillouin output exhibits randomly distributed trains of periodic pulses. Partial stabilization of amplitude fluctuations is achieved as laser pump power approaches maximum value. These experimental features are shown in time domain in Fig. 2 (a) and (b).

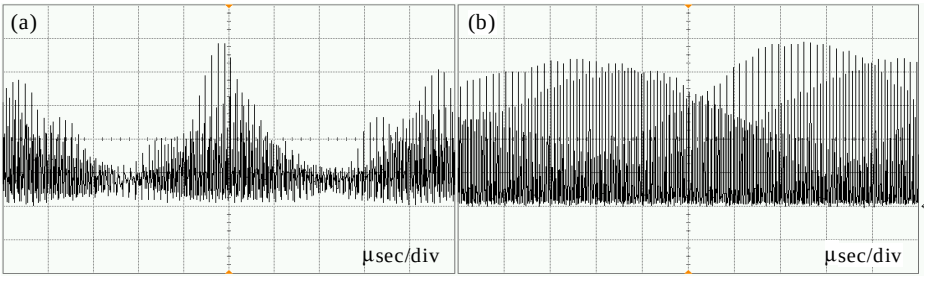


Fig. 2. Temporal structures of sBs instability (a) immediately above threshold, (b) high above threshold with chaos. The whole time scale is used in 100 µsec/division.

In the data presented, mechanical vibrations could be partially responsible for these Brillouin-temporal instabilities, because small amplitude fluctuations with similar frequencies were observed below the Brillouin threshold. The results attribute these Brillouin instabilities to phase fluctuations between direct and coupled pump intensity in the optical fiber systems.

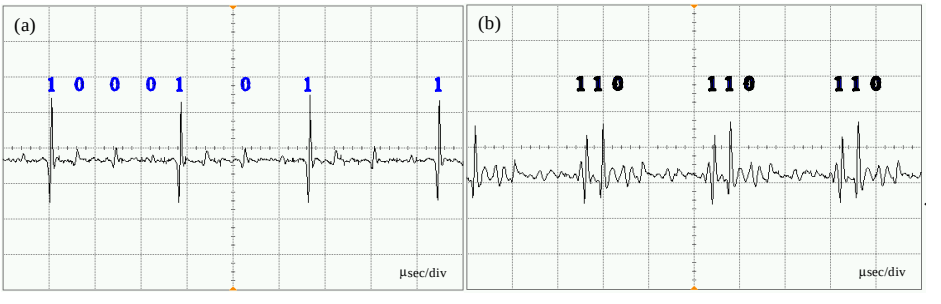


Fig. 3. Transiently controlled sBs chaos induced instabilities at immediately above threshold (a) high above threshold (b). The examples of sequence of suppression are assigned by ‘0’ and ‘1’ symbols.

The temporal evolution immediately above threshold is periodic and at lower intensities can become chaotic. If suppressing by attractor proves to control chaos then, suppressing under natural chaos can be exploited as a means of sensing structural chaos. The examples of sequence of suppression are assigned by ‘low level’ and ‘high level’ states. Multi-stable periodic states, as shown in Figure 3 (a) and (b), can lead to logic ‘0’ or ‘1’ and can in principle create large memory capacity as input bit streams in transmission data network systems. The examples of sequence of suppression are assigned by ‘low level’ and ‘high level’ states as shown in Figure 4 (a) and (b). Its implementation still requires much engineering improvements, such as arriving at a spatial resolution that is comparable to the references or speckle, and suppression of its tendency to chaos.

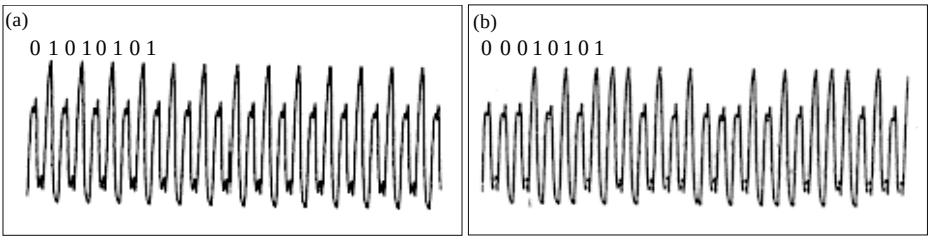


Fig. 4. Theoretically possible sequences of bifurcation are assigned by low level, high level symbols as (a) 0101010 ⋯ and (b) 0001010 ⋯

3 Conclusions

We studied that the neuron operation with steady and relaxation oscillation using controlled chaotic instabilities in optical system leads to neural networks with multistable periodic states. It can be used for optical logic and computations, optical sensing and optical arithmetic render such a scheme as the neural networks in smart structures. It is theoretically possible to apply the multi-stability regimes as an optical memory device for encoding, decoding data transmission in optical systems.

Acknowledgement

This work has been supported by KESRI (R-2005-7-047), which is funded by MOCIE(Ministry of commerce, industry and energy).

References

1. Gregory D. Vanwiggeren, Roy, R.: Communication with Chaotic Lasers. SCIENCE. **279** (1998) 1198-1200.
2. Hotate, K., Sean, S., L, Ong.: Distributed Dynamic Strain Measurement Using a Correlation-Based Brillouin Sensing System. IEEE. Photonics Letters. **15** (2) (2003) 272-274.
3. Bernini, R., Minardo, A., Zeni. L.: Stimulated Brillouin Scattering Frequency-Domain Analysis in a Single-mode Optical Fiber for Distributed Sensing. Optics Letters. **29** (17) (2004) 1977-1979.
4. Harrison, R. G., Yu, D., Lu, W., and Ripley, P. M.: Chaotic Stimulated Brillouin Scattering Theory and Experiment. Physica D., **86** (1995) 182-188.
5. Yu, D., Lu, W., and Harrison, R. G.: Physical Origin of Dynamical Stimulated Brillouin Scattering in Optical Fibers with Feedback. Physical Review A, **51** (1) (1995) 669-674.
6. Yong K. Kim, S. Lim, H. Kim, S. Oh, and Chung Y.: Implementation of Brillouin-Active Fiber Based Neural Network in Smart Structures. LNCS **3498** (2005) 987-991.
7. Yong K. Kim, J. Kim, S. Lim, and D. Kim.: Neuron Operation Using Controlled Chaotic Instabilities in Brillouin-Active Fiber based Neural Network. LNCS **3612** (2005) 1045-1051.

A Wireless Miniature Device for Neural Stimulating and Recording in Small Animals

Weiguo Song[1], Yongling Wang[1], Jie Chai[2],
Qiang Li[2], Kui Yuan[1], and Taizhen Han[2]

[1] Institute of Automation, No.95 Zhongguancun East Road
Beijing, China 100080
{kui.yuan, weiguo.song}@ia.ac.cn
[2] Department of Physiology, Xi'an Jiaotong University
Xi'an, China 710061

Abstract. This paper presents a miniature device, which is used for stimulating to and recording from the brain of an animal, based on system on chip (nRF24E1). Its performance is validated by *in vivo* experiments, in which rats are trained to run down a maze to respond to auditory instruction cues by turning right or left in order to get an electrical stimulation 'virtual reward', and by comparing spikes recorded from the brain of rats between our device and a commercially available device (Spike2, Cambridge Electronic Design Ltd.). Results show that our device can work reliably and stably, and with notable characteristics of light weight (9g, without battery), simplicity and practicality.

1 Introduction

Electrical stimulation has been widely used in electrophysiology researches and for treating a variety of neurologic disorders [1,2], and recently for the locomotion control of animals [3-5]. The neural signal recording, on the other hand, has been used to understand the functions of the brain [6-8]. Their combination would not only have significance for biomedical, artificial intelligence and cognitive researches, but also have practical value in "animal robots" [9,10].

But the commonly used electrical stimulators or neural activity recorders need either high operating voltage or cable connection, which makes them very complex and heavy, and thus, they are limited to enclosure environments or used for big shaped animals or human body [11-13]. With the advance of microelectronic technology and the validity of direct brain stimulation by digital signal, there emerged different types of miniature or implantable devices for neural stimulation or recording [3-5,11-15]. But, either they are fabricated with specific technology such as ASIC, which means they are costly and not flexible, or need complex surgery to be implanted. Although there exist some mini-telemetry devices, which are designed by discrete analogy circuit for RF transmitting and used for small animals or even insects [14,15], none of them can communicate in bi-direction. On the other hand, except for electrical stimuli, physical stimuli (such as sound) may also play a key role in behavioral experiments, and this lacks in present miniature devices. In [16], we

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 884 – 893, 2006.

presented a remote controlled multi-mode stimulator but without the function of neural activity recording. This purpose of paper is to present a miniature device for bi-direction communication, that is, it can not only generate electrical stimulation to and record the neural signal from the brain, but also give audio stimulation.

The arrangement of this paper is as follows: In section 2, the whole system design is presented, and then, preliminary performance tests are given in section 3. Finally, section 4 details the conclusions.

2 System Overview

The system is composed of two parts: the main one, which is referred to as front-end, is a miniature device carried on the back of an animal, and the other part (back-end) is a PC and a RF transceiver, which links the PC and the front-end, as shown in Fig.1. After being set from the PC, the commanding parameters are transmitted via serial and RF link to the front-end, and then, it would generate corresponding stimulating pulses to the animal brain or drive a buzzer to give sound. The neural signals from the recording electrodes, after being amplified, digitalized and processed, would be transmitted back to the PC via the same RF and serial link. The behavior patterns of the animal can be tracked and recorded automatically by a vision tracking system.

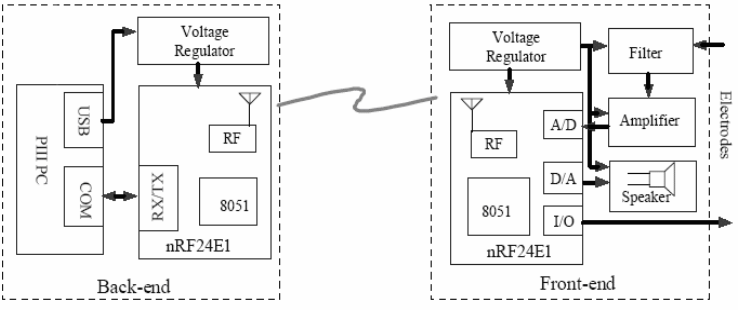


Fig. 1. Schematic of the telemetry system

2.1 Front-End

Electrodes. The stimulating electrode is self-made concentric needle by inserting Teflon coated stainless steel with diameter of 50um in a cannula with diameter of 80um, and the Teflon coating is removed from the tip for about 0.5mm. The recording electrode and the micro-drive were provided by Ma's Lab [13]. The impedances of the stimulating and recording electrode (with brain tissue) are measured to be around 80–120K Ω and 2.0–2.5M Ω, respectively. Thus, the effective stimulating current intensity delivered to animal brain is around 33 μA for 3.3V pulse trains.

Front-end Amplifiers and Filters. The front-end analog part, which is fabricated in a separated printed circuit board, is composed of amplifiers and filters, as shown in

Fig.2. The preamplifier is a differential instrumental amplifier with high input impedance (INA118). The gain of the preamplifier is fixed at 10 in order to get high bandwidth (>100 KHz, at -3dB). Before the neural signals are sent to the A/D converter of the nRF24E1, they are preamplified first, then, band-pass filtered (100 Hz ~ 10 KHz) by cascading a high-pass and a low-pass filter both with gain of 10. The power-line interference is further decreased by a 50Hz notch filter. Thus, the total gain of the system is 1000.

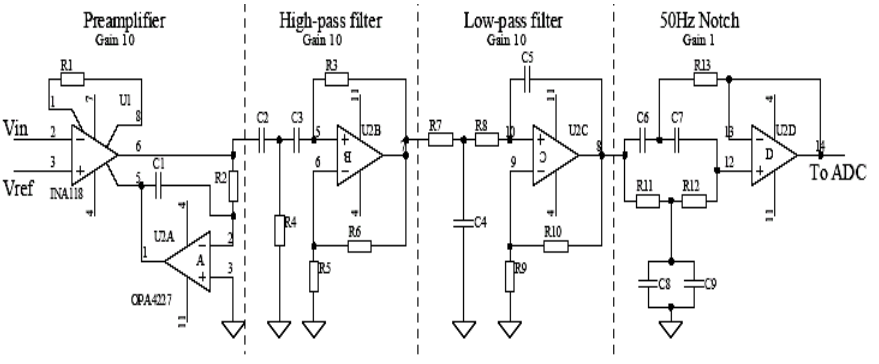


Fig. 2. Front-end analog circuit: Input is at the left; output at the right. The four sections described in the text are outlined with dashed lines. These are preamplifier with AC coupling, high-pass filter, low-pass filter, and notch filter.

Main Telemetric Recorder and Stimulator. The nRF24E1 [17], embedded with RF transceiver and 8051 microcontroller, is chosen as the main component of the miniature device for its efficient and convenient communication via serial and RF link. Also, it provides a PWM port and nine (up to) 12 bit of 100Kspls/s ADCs. As the typical frequency band of neural signal is between 100Hz and 3 kHz, the A/D conversion needs 8 bit resolution with sampling frequency of 8 kHz. In order to minimize the weigh and size of the device, all components are chosen as surface

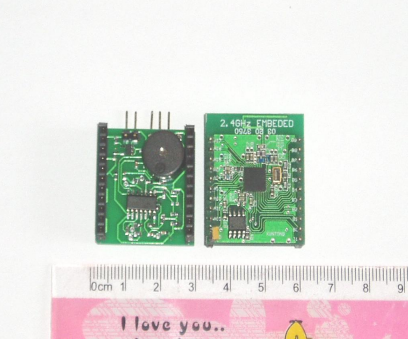


Fig. 3. The real image of the front-end analog part (left) and the main telemetry (right)

mounted package, thus the total weight is no more than 9g (without battery). The real image of the front-end is shown as in Fig.3.

Stimulating pulses can be set either monopolar or bipolar. By setting the digital I/O ports (P0.5, P0.6) of the nRF24E1 to high and low level alternatively and simultaneously, bipolar stimulating pulses could be achieved effectively without the need of negative power. The PWM port is used to drive a buzzer to generate sound with different frequency, and all the idle ports of the nRF24E1 are set to low level in order to minimize the power consumption and to minimize the noise. The miniature device decodes commands from the PC and generates corresponding stimulating pulse trains or drives a buzzer to give sound.

2.2 Back-End

RF Transceiver. The back-end RF transceiver acts as a bridge across the host PC and the front-end via the serial and RF channel (check Fig.1 for clarity), consecutively. It is an adapted version of the front-end main telemetry by adding serial data link, that is, a MAX202C is used for the voltage conversion between the MCU and the PC serial port. Also it can be replaced by a USB interface to achieve higher transfer rate.

Serial Communication Protocol. The baud rate of the serial communication between the PC and the back-end is set 19.6Kbps. Thus, the time delay of the commanding package is less than 5ms, which can meet the experiment requirements. The protocol for transmit from back-end to front-end is as shown in Table.1, where two bytes of synchronization head (0xFF, 0XAA, which show noise resistance) are followed by five bytes of commanding package (DATA). For transmitting to the front-end, they represent stimulating polarity (monopolar/bipolar), pulse width (50us~10ms), pulse number (1~255), train interval (1s~255s), and train frequency (0.1Hz~255Hz), respectively. The last byte (CHECKSUM) is the check sum of the whole data package. But, for the received data package(neural signal), two bytes of synchronization head (0XAA,0xFF) is followed by two bytes of neural spiking data (DATA), which contains the number of spikes detected in a given time.

Table 1. Serial communication protocol

Byte No.	1	2	3-7	8
Data	0xFF	0xAA	DATA	CHECKSUM

2.3 RF Link Between the Front-End and the Back-End

RF Configuration Words. The data packet for RF link is shown in Table 2. The PREAMBLE is the package head, which is added and removed automatically at the transmitter and receiver respectively. The ADDRESS is a receiving address, which needs to set at the transmitter and would be removed automatically at the receiver; the DATA is the payload, the maximum width (bits) of which is 256 minus the address width (bits) and the check width (bits); the CRC represents a 8 or 16 bit checksum. First byte of the receiving address should not start with 0x55 or 0xAA, which might

be interpreted as part of preamble and cause address mismatch for the rest of the address. In general, more bits of address and checksum will give less false detection, and thus lower data packet loss, but it would result in low transmit efficiency. In order to communicate reliably and with short delay, it is configured as single receive channel, 40 bits receive address width, 16 bits check width and 6 bytes of payload. As utilizing 250 kbps instead of 1Mbps would improve the receiver sensitivity by10 dB, we set lower rate to achieve high sensitivity.

Table 2. Radio frequency (RF) configuration works

Byte No.	1	2-6	7-11	12-13
Data	PREAMBLE	ADDR	DATA	CRC

RF Bi-direction Communication Protocol. In order to establish a bi-direction RF communication at a transferring rate of 64kbit/s, nRF24E1 should transmit and receive data package alternatively in the right timeslot, as shown in Fig.4. In each timeslot the A/D result must be read out and the PWM updated. Synchronization of the master and slave is achieved by coupling the activities on air to a counter. The master always transmits in timeslot 5 to 7, and once the slave has received a valid package, the RF end is turned off and the slave counter is reset to 19. Each package transmits every 3 ms.

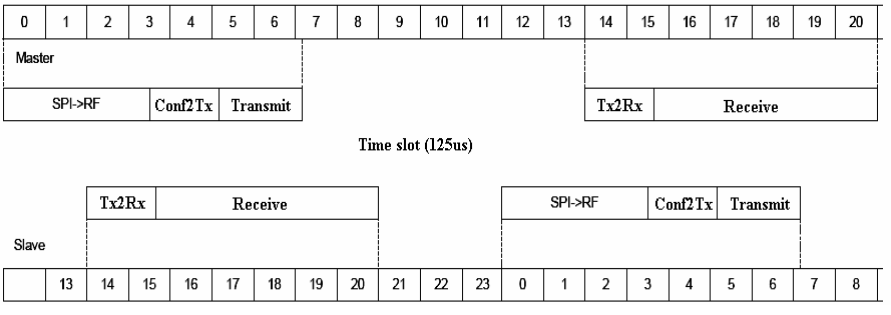


Fig. 4. Timeslot division for full-duplex link

2.4 Power Source

The power supply of the whole system is composed of two parts independently: The power at the back-end is derived from the USB port of the host PC by a voltage regulator (LM2576) from 5V to 3.3V, which is very convenient and practical especially in open field. A rechargeable polymer-battery (3.7V, 160mAh, 5g) is chosen for the front-end power source, after regulated by two low dropout voltage regulators (MAX884), they are used for the analog circuit and the main digital MCU (nRF24E1), respectively. The analog ground is derived from the positive rail by a voltage divider and a unity gain follower. The follower's output is served as the

reference to the experimental subject. The battery is clamped between the two PCBs of the front-end, and it can supply the system for 10 hours continuously.

3 Methods

3.1 Animal Preparation

Three adult male SD rats, weighing 290~320g, were used. The surgery was conducted under the approval of the Animal Experimental Committee of XJTU. Subjects were stereotaxically implanted with stimulating microelectrodes in the MFB (L:-7.5mm; A: 2.6mm; P: 2mm) and recording electrodes in the striatum (L: -2.8mm; A: 1.7mm; P: 7.8mm) with reference to [18], and then four stainless steel screws were driven in the skull around the cannula evenly. After the surface of the skull was cleaned of fascias and thoroughly dried, the screws and the cannula were secured to the skull firmly with dental acrylic. Finally, the animals were injected with antibiotics for recovery in the following three days.

3.2 Experiments for Stimulation Performance

A Y-maze paradigm was used for behavioral training, as described in [16]: Rats were trained to run down the maze to respond to auditory instruction cues by turning right or left to receive electrical stimulating reward (MFB) for five days. A daily training session consisted of two 5min sessions with ten minutes break in between (each tone presented randomly). Whenever the maze presented the rats a choice of turning left or right, the micro-stimulator generated a specific sound. If the rat made the corresponding turn, it was immediately given electrical stimulation in the MFB and recorded as a ‘correct’, or else, it was deprived of electrical stimulation and recorded as a ‘fault’. The animal’s behavior was recorded by a video tracking system, and the number of spikes detected from recording electrodes is sent back to PC simultaneously. Parameters used for MFB stimulating were as follows: biphasic, amplitude 3.3V, duration 0.5ms, frequency 100Hz, pulse number 10; the parameters of the sound, used as instruction cues, were as follows: one 1min of 1KHz for left turn and one 1min of 5KHz for right turn.

3.3 Spike Recording and Detecting

Although many kinds of spike detection algorithms have been proposed [19], most of them are not robust enough or time consuming, and thus, they can only be used for off-line analysis. Recently, the nonlinear energy operator (NEO), which represents the instantaneous energy of a signal, has been used for spike detection and has the characteristic of simplicity and low computational burden [20,21]. Thus, we base our detection on it. The discrete form of the NEO is

$$\Psi[x(n)] = x(n)^2 - x(n+1)\cdot x(n-1). \quad (1)$$

4 Results

4.1 Stimulation Performance

The stimulation performance of the device is validated by sending commands from a distance up to 200m away. Fig.5 shows the stimulating train of biphasic pulses delivered to the stimulation channel. The device can deliver arbitrarily specified stimulus trains according to the commands sent from the back-end host PC with high fidelity. Statistical results from rat conditional reflex training, as in Fig. 6, shows that in an average of five days rats reached the 93.5% correct criterion for acquisition and the averaged correct turns increase by three times. The preference of the animal to the MFB stimulating also shows the effectiveness of the device.

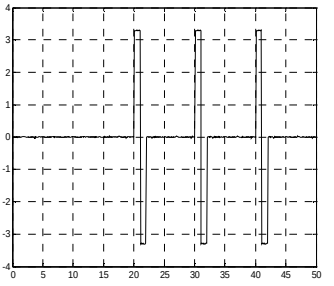


Fig. 5. The output biphasic stimulation waveform (amplitude: 3.3V, pulse number: 3, duration: 1 ms, and frequency: 100 Hz) taken from oscilloscope, when commands sent 200m away

4.2 Recording Performance

The recording performance of the front-end is compared by simultaneously recording spikes from electrodes with our device and with a commercially available device (Spike2, Cambridge Electronic Design Ltd.), as shown in Fig.7. Both detect almost

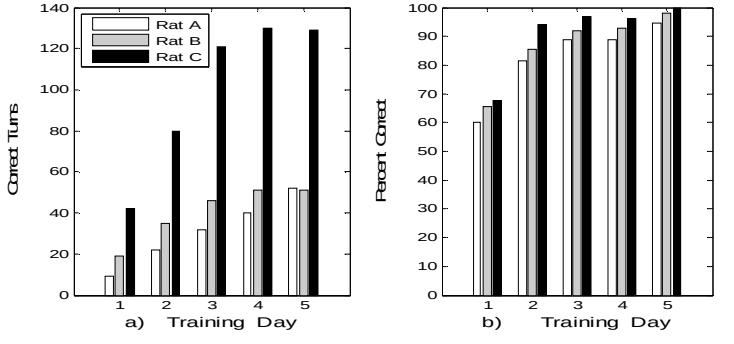


Fig. 6. The training process for three rats. In an average of five days rats reached the 95% correct criterion for acquisition b) and the averaged correct turns increase by three times a).

the same occurrence time for each spike, although their shapes are different. Fig.8 shows the raw data (upper), the output of NEO (middle) and the detected results (black dots in lower), which represents the firing frequency of a neuron and would be transmitted back to the PC in real time for further analysis.

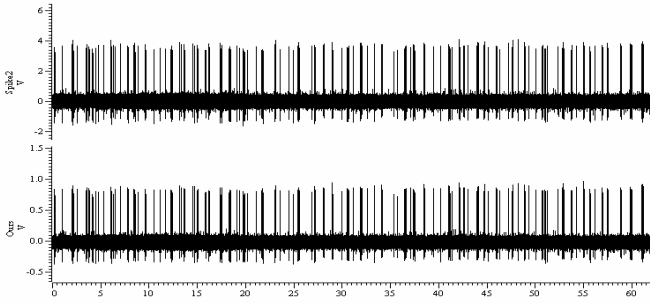


Fig. 7. The recorded neural signals coming from the same electrode implanted in the striatum of a rat but from two channels (top: Spike2; bottom: our device)

5 Discussions

This paper presents a miniature device, which not only can generate stimulations (audio or electrical), but also can record neural activities. The electrical stimulating pulses can be monopolar or bipolar, and the audio stimuli with different characteristics. The effectiveness and performance of this system were validated from experiments. The results showed that it can not only provide a platform in behavioral and neuroscience research on behaving small animals, but also can be used as an efficient animal-machine-interface and even for "animal robot". But, there leaves

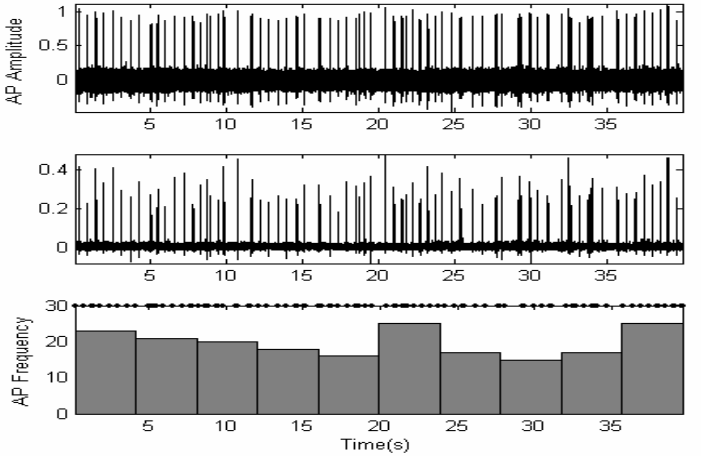


Fig. 8. Recorded raw data, the output of NEO, and the detected spikes (black dots)

more to be improved, such as the stimulation channels, stimulation intensity, and integration of more robust detection algorithms.

Presently, we have chronically implemented electrodes in the brain of the rats, and are trying to explore the relationships among electrical stimulation, spiking characteristics and behavioral responses [6,7]. Still further, we hope it could form a closed loop for the brain-machine interface (BMI) [8,9].

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 60375026).

References

1. Cavuoto, J.: Neural Engineering's Image Problem. IEEE Spectrum. April (2004) 32-37
2. Mogilner, A.Y., Benabid, A.L., Rezai, A.R.: Brain Stimulation: Current Applications and Future Prospects. Thalamus Relat. Syst. 1 (2001) 255-267
3. Holzer, R, Shimoyama, I.: Locomotion Control of a Bio-Robotic System via Electric Stimulation. In. Proc. of the IEEE/RSJ Int. Conf. 3 (1997) 1514-1519
4. Talwar, S.K., Xu, S.H., Hawley, E.S., Weiss, S.A., Chapin, J.K., Moxon, K.A.: Behavioural Neuroscience: Rat Navigation Guided by Remote Control. Nature. 417 (2002) 37-38
5. http://www.newscientisttech.com/channel/tech/mg18925416.300.html
6. Romo, R., Hernandez, A., Zainos, A., Lemus, L., Brody, C.D.: Neuronal Correlates of Decision-Making in Secondary Somatosensory Cortex. Nature Neurosci. Vol.5. No.11 (2002) 1217-25
7. Nicolelis, M.A.L.: Actions from Thoughts. Nature. 409 (2001) 403-407
8. Chapin, J.K., Moxon, K.A., Markowitz, R.S., Nicolelis, M.A.: Real-time Control of a Robot Arm Using Simultaneously Recorded Neurons in the Motor Cortex. Nature Neurosci. Vol.2. No.7 (1999) 664-670
9. Mussa-lvaldi, F.A., Miller, L.E.: Brain-machine Interfaces: Computational Demands and Clinical Needs Meet Basic Neuroscience. Trends in Neurosci. Vol.26. No.6 (2003) 329-334
10. http://www.newscientist.com/article.ns?id=dn6429
11. Arabi, K., Sawan, M.A.: Electronic Design of a Multichannel Programmable Implant for Neuromuscular Electrical Stimulation. IEEE Trans. Rehab. Eng. Vol.7. No.2 (1999) 204-214
12. Obeid, I., Morizio, J.C., Moxon, K.A. et al.: Two Multichannel Integrated Circuits for Neural Recording and Signal Processing. IEEE Tran. Biomed. Eng. Vol.50. No.2 (2003) 255-258
13. Lei, Y.L., Sun, N.L., Wilson, F.A.W., et al.: Telemetric Recordings of Single Neuron Activity and Visual Scenes in Monkeys Walking in an Open Field. J. Neurosci. Methods. Vol.135. (2004) 35-41
14. Nieder. A.: Miniature Stereo Radio Transmitter for Simultaneous Recording of Multiple Single-Neuron Signals from Behaving Owls. J. Neurosci. Methods. Vol.10. (2000) 157–164
15. Takeuchi, S., Shimoyama, I.: A Radio-Telemetry System with Shape Memory Alloy Microelectrode for Neural Recording of Freely Moving Insects. IEEE Tran. Biomed. Eng. Vol.51. No.1 (2004) 133-137
16. Song, W.G., Chai, J., Han, T.Z., Yuan, K.: A Remote Controlled Multimode Micro-Stimulator for Freely Moving Animals. Acta Physiologica Sinica. Vol.58. No.2 (2006) 183-188

17. http://www.nordic.com
18. Bao, X.M., Shu, S.Y.: The Stereotaxic Atlas of the Rat Brain. People's Sanitation Press, Beijing (1991)
19. Lewicki, M.S.: A Review of Methods for Spike Sorting: The Detection and Classification of Neural Action Potentials. Network: Computation Neural System. 9 (1998) R53-R78
20. Kim K.H., Kim S.J.: Neural Spike Sorting Under Nearly 0-dB Signal-to-Noise Ratio Using Nonlinear Energy Operator and Artificial Neural-Network Classifier. IEEE Tran. Biomed. Eng. Vol.47. No.10 (2000) 1406-1411
21. Mukhopadhyay, S., Ray, G.C.: A New Interpretation of Nonlinear Energy Operator and Its Efficacy in Spike Detection. IEEE Tran. Biomed. Eng. Vol.45. No.2 (1998) 180-187

An SoC System for the Image Grabber Capable of 2D Scanning

Cheol-Hong Moon[1] and Sung-Oh Kim[2]

[1] Gwangju University, Gwangju, Korea
chmoon@gwangju.ac.kr
http://web2.gwangju.ac.kr/ chmoon/
[2] Gwangju University, Gwangju, Korea
Countrykims@nate.com

Abstract. This paper reports the implementation of an SoC (System on a Chip) IP(Intellectual Property) design for an Image Grabber that is capable of 2D scanning. The vast improvements in integrated circuit technology and semiconductor industry have allowed complicated hardware systems to be implemented into a single Chip. Previously, hardware designed by external buses was slow. However, the speed of the system has been improved by peripheral logics that are designed on a single chip. In addition, both hardware and software can be constructed together resulting in a significant decrease in research and development time. Four IPs were created in this study. I2C IP was developed for the Image Decoder. The NTSC analog signal from a CCD camera was converted to YUV422 using an image decoder and stored in a frame buffer by IMAGE IP. TFT-LCD IP was developed to display the data on an LCD. A robot IP was developed to control a perpendicular coordinate robot. Finally, system level software and PC software was developed to examine the SoC system.

Keywords: SoC, IMAGE IP, Perpendicular Coordinate Robot IP, TFT-LCD IP.

1 Introduction

SoC is not only hardware logic but also a Processor, ROM, RAM, Controller, and Peripheral Device Circuit that is contained in a single Chip. Previously, complicated circuits were made to implement an image grabber capable of 2D scanning. These complicated circuits were converted to hardware IPs and transformed onto a single Chip.[1] In addition, hardware and software was designed at the same time and verified. This process shortened the development time. The NTSC analog signal from a CCD camera was transformed into YUV422 format using an image decoder. The transformed YUV422 data was stored in a frame buffer using Image IP. The stored data in the frame buffer was read by ARM922T. The data was then transmitted by a DMA controller to a TFT-LCD driver and displayed on a TFT-LCD. In addition, it converts the data into to x and y coordinates, captures the images and measures the movement distance by encoders using Perpendicular Coordinate Robot IP.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 894–903, 2006.

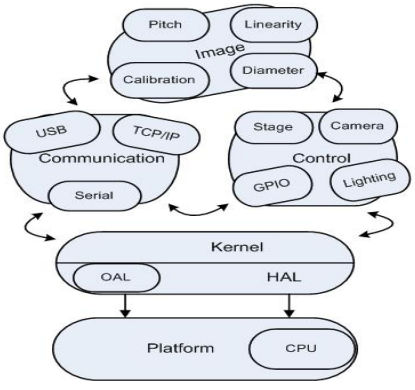


Fig. 1. An SoC System Platform

2 Hardware IP Design

2.1 SoC System

The SoC-Chip used in this paper was Excalibur, which was purchased from the ALTERA company. This Chip contains an ARM922T Processor, a hundred thousand Gate Programmable Logic, Interrupt Controller, Timer, Memory Controller, DPRAM, SRAM, and PLL etc. In the inner parts, each bus is connected to an AMBA bus.[2][3] Figure 2 shows a block diagram of the Excalibur chip used to implement the IPs.

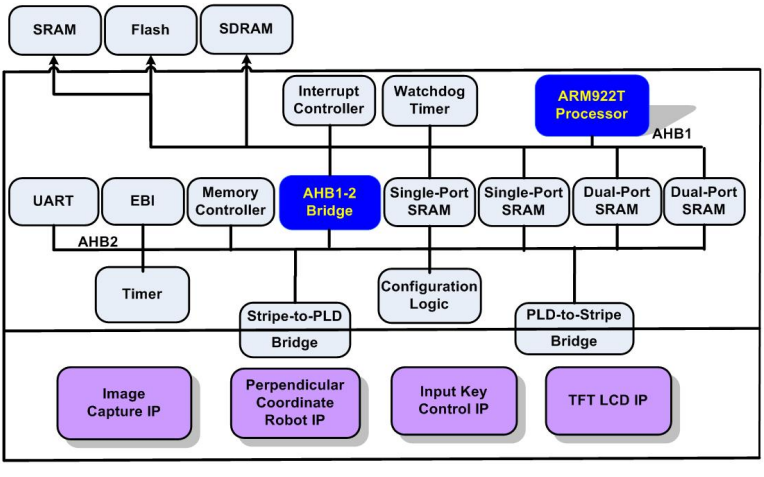


Fig. 2. Excalibur Block Diagram and developed IPs

2.2 SoC Hardware

Figure 3 shows the Logic Block Diagram examined in this paper. An SoC system consists of the following five blocks to drive the two motors and encoders. ARM922T CPU Logic Block, AMBA-Bus Multi Slave Block, TFT-LCD DMA-Controller and driver logic, IMAGE IP and Perpendicular coordinate Robot IP.

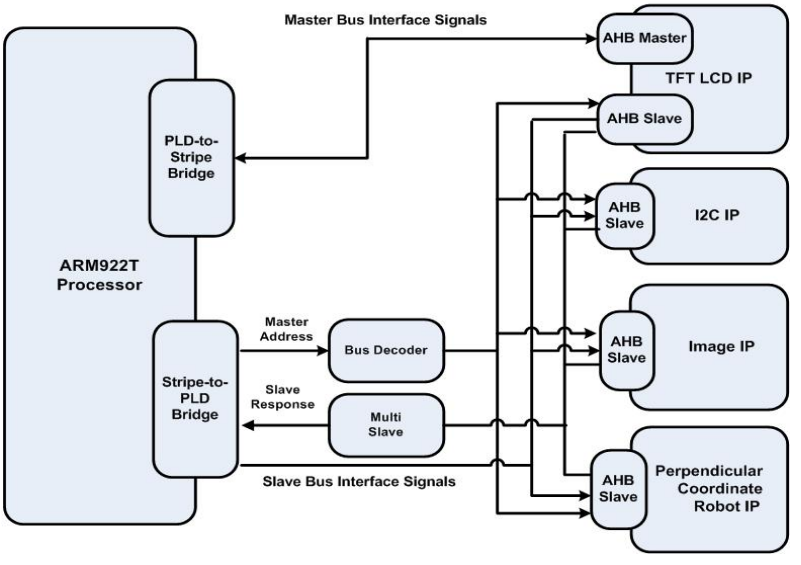


Fig. 3. Hardware Logic Block Diagram

2.3 IMAGE IP

The image decoder Chip converts the NTSC analog signal originating from a CCD image sensor to YUV422 format. A frame buffer is used to store and convert one field of image data.

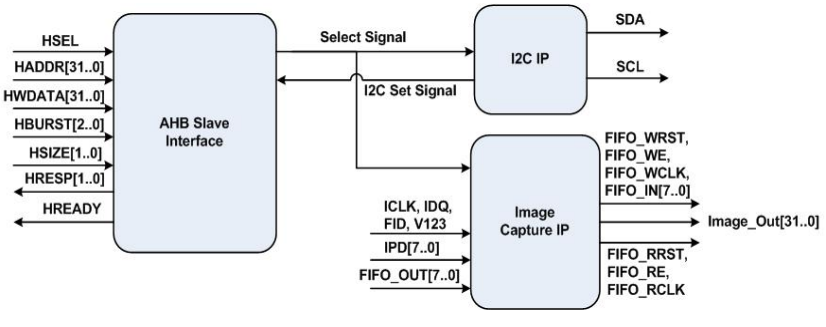


Fig. 4. IMAGE IP

IMAGE IP Hardware. Figure 4 shows the IMAGE IP. The IMAGE IP consists of 3 Logic Blocks. An AHB Slave Interface block controls the flow of data. The I2C IP block implements the logic block, which can set the I2C of the Image Decoder.[4][5] The Image Capture IP block implements the logic block that can capture one field of image data.

IMAGE IP Block Diagram. Figure 5 shows the IMAGE IP Logic Block Diagram. The IMAGE IP Logic Block consists of 4 Logic Blocks.

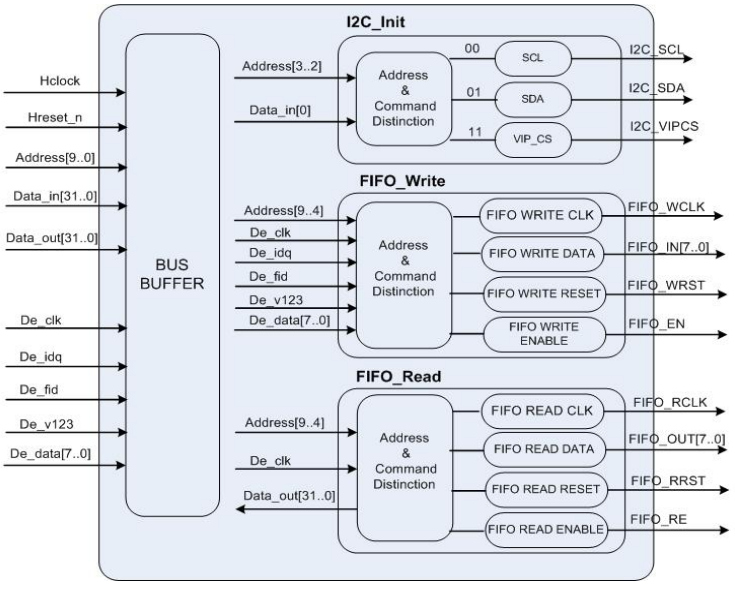


Fig. 5. IMAGE IP Logic Block Diagram

2.4 Perpendicular Coordinate Robot IP

The Perpendicular Coordinate Robot is organized by two stepping motors corresponding to the X and Y axis. In the upper axis, a CCD camera is set to measure the display element. Each stepping motor has been given an encoder to measure the moving distance. The encoder can measure both directions. Its rotation pulse is 1000. Coordinate Robot IP was designed to support these hardware specifications.

Hardware of Perpendicular Coordinate Robot IP. The AHB Slave Interface block shown in Figure 6 is a Slave-interface of the AMBA bus that holds the Perpendicular Coordinate IP in one Slave. The Motor Controller block is the x and y coordinates that measures the movement distance and controls the stepping motor containing the two-encoder logic.

Perpendicular Coordinate Robot IP Block Diagram. Figure 6 shows the Perpendicular coordinate Robot IP Logic Block Diagram.

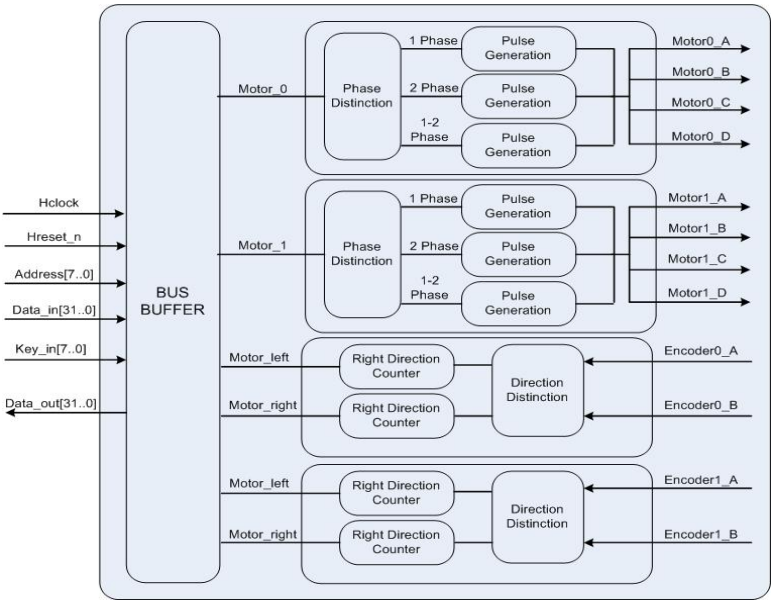


Fig. 6. Perpendicular coordinate Robot IP Logic Block Diagram

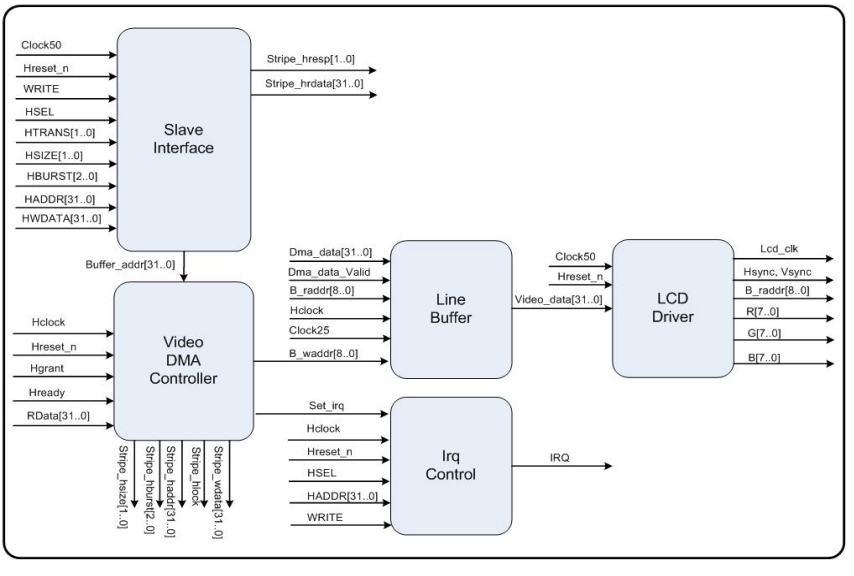


Fig. 7. TFT-LCD IP Logic Block Diagram

2.5 TFT-LCD IP

The Image Data stored in the Frame Buffer is transmitted to the TFT-LCD driver by the DMA Controller. The transmitted data is converted to 240 * 320 TFT-LCD, and is displayed on the TFT-LCD.

TFT-LCD IP Hardware. Figure 7 shows the TFT-LCD IP Logic Block Diagram. The video DMA Controller block shown in Figure 7 is the DMA Controller, which transmits the data read from the frame buffer to the DMA and transmit this data to the TFT-LCD driver.[6] The Slave Interface block is a Slave-interface of the AMBA bus and the LCD driver block is a TFT-LCD driver.

3 Experiment Result

3.1 Perpendicular Coordinate Robot IP Simulation

Figure 8 shows the simulation for the stepping motor x and y coordinates and two encoders. First of all, the simulation contains Clk, Reset n, Address, Data in signals to control the stepping motors. The simulation receives the Key in value to control the left, right, and acceleration. The stepping motor signals contains Motor1 A, B, C, D and Motor2 A, B, C, D. The full-mode method was used to drive the motor. The encoder signals used Encoder1 A and B, and Encoder2 A and B. The data out signals were used to transmit the distance value to the ARM922T.

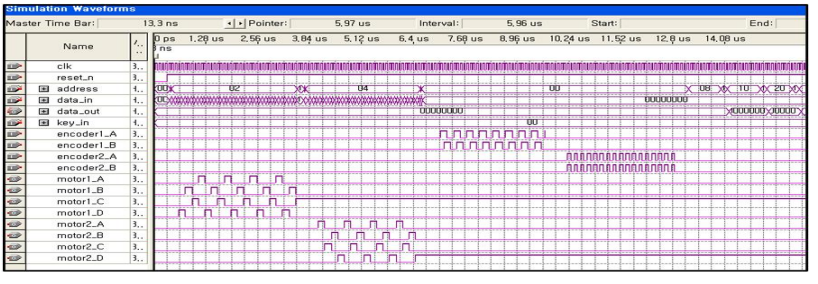


Fig. 8. Perpendicular Coordinate Robot IP Simulation

3.2 TFT-LCD IP Simulation

Figure 9 shows a simulation of the Video Driver. There are Clock25 and the Reset n signal to drive the video driver block. The Video data and Video dimension signals were used for the image data and size. R,G,B, Hsync, Vsync, Pwr_en, Clockext and the DE signal, and R.G.B of the Video Driver were used to drive the TFT-LCD.

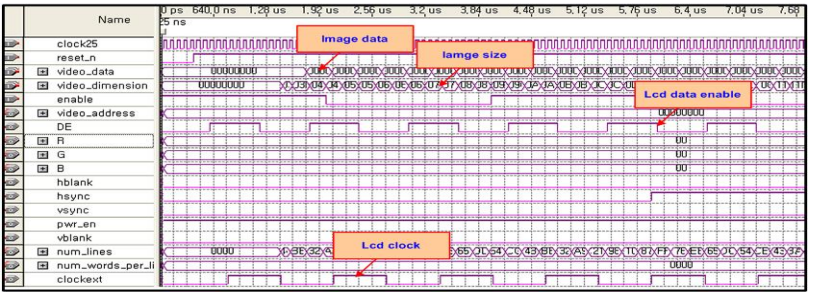


Fig. 9. Video driver Simulation

4 Experiment of an SoC System

The NTSC analog signals from the CCD camera was converted to YUV422 format using the image decoder chip. The converted YUV422 data was saved as frames in FIFO. The IMAGE IP was developed to convert the data to RGB data through an ARM922T processor. The converted RGB data was displayed on a TFT-LCD through a DMA controller. The ARM922T transmitted the YUV422 image data to debug the PC. A Debug PC controlled a Perpendicular Coordinate Robot using Perpendicular Coordinate Robot IP. The experiment was carried out to display the image data received from the ARM922T on a PC screen.

Experimental Environments. In this experiment, a system composed of a debug PC, an SoC system, a perpendicular coordinate robot and a display element was configured. Figure 10 shows the system flow. A CCD camera was used to capture the image data of the target. The captured analog data was processed by SoC IPs and system level software. A frame of the image data was formatted into YUV422 format and saved in RAM. The saved image data was then transmitted to a debug PC. A Debug PC was used to convert the YUV signal to RGB data and then to a device independent bitmap, which was displayed on a screen. A Debug PC and key entry was used to control the Perpendicular Coordinate Robot.

Experiment Methodology. An SoC system experiment was classified as the IP logic part, Firmware level software and PC software. The function control in image I/O was performed using PC software and communication. A Perpendicular Coordinate Robot was controlled through communication with PC software or by key entry from the outside. Figure 11 shows the image I/O process between the SoC system and the debug PC. The left block of Figure 11 shows the SoC IP logic part where the actual I2C IP, IMAGE IP, and TFT-LCD IP are located, the middle block SoC S/W is the system level software of the ARM922T. Finally, the right block is PC software in debug mode to display the image data requests and a CCD camera that takes images on a PC. Figure 12 shows the perpendicular coordinate robot control process between the SoC system and debug the PC . The left block in Figure 12 is the PC software and gives a perpendicular

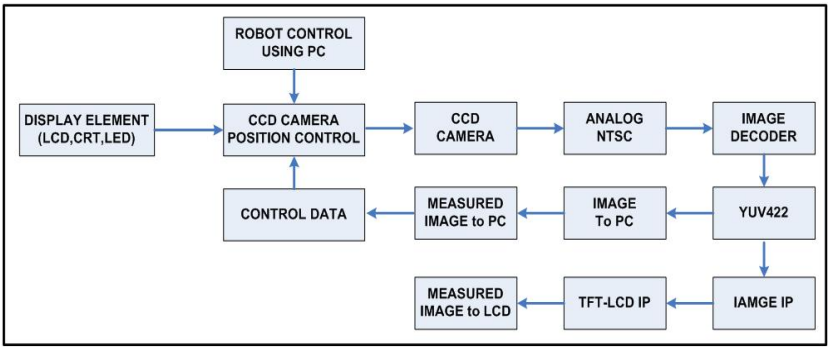


Fig. 10. The system control flow

coordinate operation command to the SoC S/W. The middle block is SoC IP and the system level software of ARM922T, which receives the command from the PC software or entry key, controls a perpendicular coordinate robot. Finally, the right block is the hardware part, which connects the keypad to an SoC system to give a key value, and allows the stepping motor driver and rotation pulse signals of the encoder.

Experiment Result. These experiments were designed to verify the simulation of each IP and to check if they operate according to the simulation result when they ported to an actual SoC. The results confirmed that the simulation of each

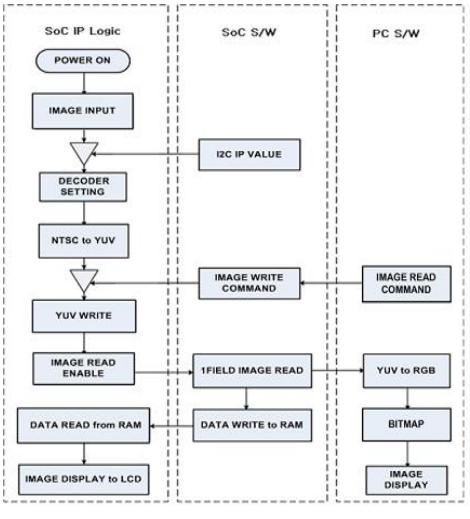


Fig. 11. Input/output of image

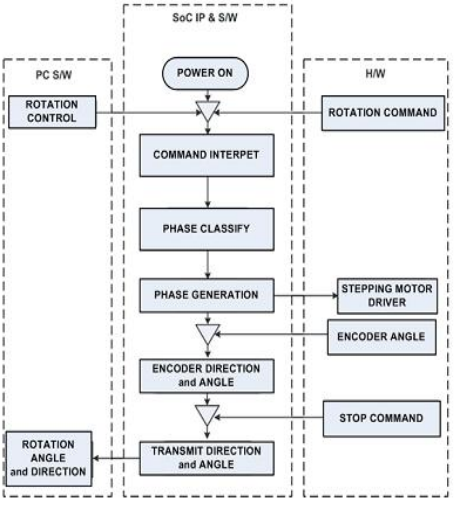


Fig. 12. The Perpendicular Coordinate Robot control

IP would operate according to the synchronization signals. However, when they were ported to the actual SoC, the result was different from the simulation. In addition, when acquiring the image of the display elements with a CCD camera, the color and vividness became dull due to an error in abstracting the unverified image data and the focus distance of the CCD camera. However, the SoC IPs were developed to measure the display element. They contain a hardware system IP, which used to be very complicated and slow, and IPs that can be reused. In addition, it was implemented into a single chip SoC to simplify the hardware system and to obtain a faster operation result.

The following Figure 13 and 14 show SoC system, PC, Perpendicular Coordinate Robot,Program of Perpendicular Coordinate Robot used in this experiment.

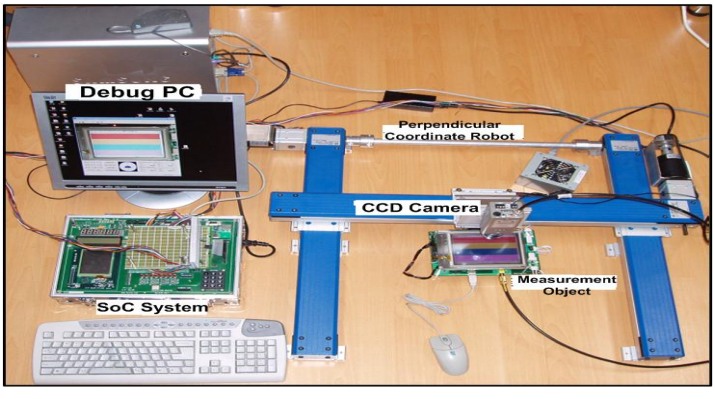


Fig. 13. System to measure LCD

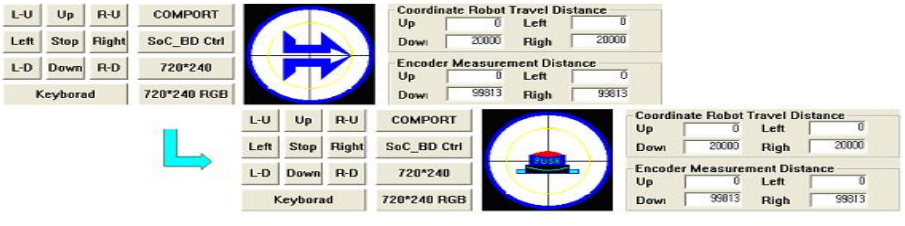


Fig. 14. Program of Perpendicular Coordinate Robot

5 Conclusion

IMAGE IP, perpendicular coordinate robot IP and TFT-LCD IP were used to measure the display elements. A faster processing speed could be obtained by replacing the complicated circuits with the SoC and implementing each device with logic. Although the devices are currently implemented as IPs, a future

system should be larger and allow multi tasking as well as a network and multi media system. This study developed the system level software only rather than the OS itself. However, This system might be applicable to many other fields if it can be applied to an OS. The Soc is expected to make a significant contribution to the development of a faster system because various systems can be condensed into a single SoC, and hardware and software can be developed together.

References

1. Huins.: SoC design education, Huins, 2003.
2. ARM.: AMBA Specification, Rev2.0, ARM, 1999
3. Altera.: Excalibur Devices Hardware reference Manual version 3.1, November 2002
4. Philips.: "The IIC Bus Specification Ver 2.1", Jan,2000
5. Philips.: SAA7114H PAL/NTSC/SECAM video decoder product Specification 15, Mar, 2000
6. Samsung Electronics.: LTS350Q1-PD1 Product Data Sheet, 2003

Hardware Implementation of AES Based on Genetic Algorithm

Li Wang, Youren Wang, Rui Yao, and Zhai Zhang

College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing Jiangsu 210016, China
qinwangli_1982@tom.com,
wangyrac@nuaa.edu.cn

Abstract. The high performance cryptographic chip is the core unit of the network information safety equipments. This paper proposes the design approach of the AES cryptographic system based on reconfigurable hardware, develops the method of key-sequence generation with the genetic algorithm that realizes different cipher key in every encryption round. The system has been implemented on Virtex-E FPGA. The result proves that the new design technology is feasible, and the security level of the AES is improved.

1 Introduction

Transmission and storage of sensitive data in open networked environments is rapidly growing. Along with it, grows the need of efficient and fast data encryption. Software implementation of cryptographic algorithm cannot provide the necessary performance when large amounts of data have to be moved over high speed network links[1]. Hardware implementation of cryptographic algorithm is faster and saver. Nowadays the hardware implementations are both based on FPGA (Field Programmable Gate Array) and ASIC (Application Specific Integrated Circuits). FPGA are very well suited for high speed cryptography, as it can provide the required performance without the excessive cost of ASIC manufacturing process[2][3].

AES algorithm is a symmetric block cipher, and it processes data blocks of fixed size (128 bits) using cipher key of length 128, 196, or 256 bits [4]. This paper proposes a reconfigurable hardware implementation of AES, which is flexible and fast. The architecture of the AES implementation and the proposed key-sequence generation will be discussed. The key-sequence generation which improves high security level of the AES algorithm is based on genetic algorithm; so that the cipher key is different in every encryption round. The implementation of the AES cryptographic system will be realized on Virtex-E FPGA.

2 Reconfigurable Hardware Implementation of AES algorithm

2.1 AES Hardware System

The AES algorithm is a symmetric block cipher that is excellent in security, efficiency, and flexibility, widely applied in hardware implementation. AES performs 4

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 904 – 907, 2006.

discrete transformations—AddRoundKey, SubBytes, ShiftRows, and MixColumn. The introduction of these 4 transformations is shown in [5].

This paper combines hardware with software to realize the AES. The system has two parts which are hardware data processing and software control. The hardware is realized on RC1000 which is consisted by a FPGA from Xinlinx Virtex-E, some SRAMs (Static Random Access Memory) and interface control circuit. The hardware architecture of the cryptographic system is showed in Fig.1. Functions realized on FPGA are changed on-line to satisfy different requests and provide flexibility. The FPGA is used to complete the AES algorithm and generate the key-sequence real-time; following the hardware netlist of the AES algorithm is generated by DK. The software control is completed by Visual C++, which is used to download the bit stream generated by the ISE to FPGA and save the result.

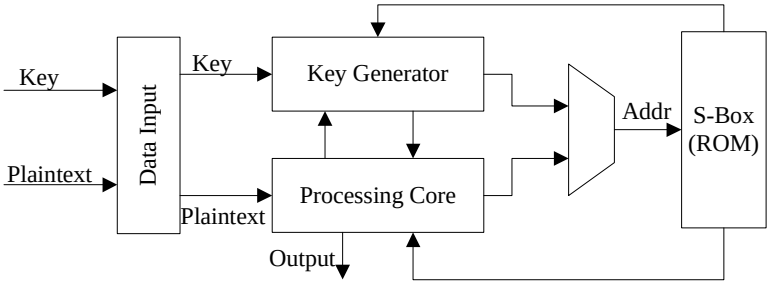


Fig. 1. Hardware Architecture of the AES Cryptographic System

2.2 The Generation of the Key-Sequence

The generation of key-sequence in AES is based on genetic algorithm, and its basic idea is using crossover, mutation operation to create the data string randomly. The 128 bits cipher key is divided into two parts which are the parents of genetic algorithm, and each part has 64 bits. The detailed implementation of the key-sequence is introduced as follows:

(1) The value composed by the zero bit and first bit of second parent represent how many times the parents should rotate right.
(2) If the second bit of the parent is 1, the first parent should bitwise not. If the third bit of the second parent is, the second parent should bitwise not.
(3) The value composed by the forth bit, fifth bit, sixth bit, and seventh bit and then added 25 represents the place where the parents crossover.
(4) The value composed by the eighth bit, ninth bit, tenth bit, eleventh bit, twelfth bit, thirteenth bit and added 1 represents the place in where the first parents mutation.
(5) The value composed by the fourteenth bit, fifteenth bit, sixteenth bit, seventeenth bit, eighteenth bit added 1 represents the place where the second parents mutation.

After the 5 steps, two new individuals are generated to build up a new 128 bits cipher key which is used in the next round encryption.

3 Results

AES-128 has been implemented on Virtex-E. This paper takes 256 as cipher key and 0 as plaintext. Table 1 shows the results of five rounds encryption.

Table 1. The Results of Five Rounds Encryption

Round	Plaintext	Cipher key	Ciphertext
1	0	0x100	0x658d61cf764b6e2d304bf9ce21da41a6
2	0	0xead55005d6127045cbd3c8cf870f188e	0x6368facdea15c2215c1b735c2dcceb1b
3	0	0x3c5c1a0cd0f78d4d0c8141ea69a1c314	0x7819e1b2740ce57276d12b2f50075640
4	0	0xe5a7b427abd7e839680e74859cb1ee8f	0x86f3bb463f485754d0c3aa30e2fca384
5	0	0x7d63d6d626d977b81ef654fa3ba80856	0xccc636d8e45654e240dd18b5849d8601

From the table 1, the plaintext is the same but the ciphertext is different because the round key created by genetic algorithm is different. The hardware crptographic implementation of AES has the advantage of AES itself and the randomness of the round key is improved, so that it can resist some cryptographic attack. The resources of FPGA can be repeatedly used in each process of the cryptographic implementation which is fast and flexible, and the availability of the FPGA resource is also improved.

The MixColumns transformation of AES mixes the bytes in each column by the multiplication of the state with a fixed polynomial module with its coefficient in GF (2^8). The method of multiplication holds a lot of resource, so shift and XOR operation are used to realize the MixColumns transformation which reduces logic resource and simplifies the circuit. The total NANDs of the AES cryptographic system are 50700.

The transfer rate to and from RC1000 can be computed. The transfer rate to RC1000 is 112.30 Mb/s and the transfer rate from RC1000 is 117.44 Mb/s.

4 Conclusion

The proposed cryptographic system of AES based on reconfigurable hardware and genetic algorithm has been realized on Virtex-E FPGA. It is flexible and has high encryption speed. The genetic algorithm improves the security level of the AES algorithm itself because key-sequence is generated in every round encryption. The cryptographic system of this paper can also resist some cryptographic attack, and also can resist the difference analysis and linear analysis.

References

1. Andreas Brokalakis, Athanasios P. Kakarountas, Costas Goutis: A High-throughput Area Efficient FPGA Implementation of AES-128 Encryption. IEEE Workshop on Signal Processing System Design and Implementation, Athens, Greece (2005) 116-12

2. Xin-miao Zhang, Keshab K. Psrhi: High-speed VLSI Architectures for the AES Algorithm. IEEE Transaction on Very Large Scale Integration (VLSI) Systems, vol.12 (2005)957-967
3. Guo-Zhen Xiao, En-Jian Bai, Xiao-Juan Liu: Some New Developments on the Cryptanalysis of AES. Acta Electronica Sinica, vol 10 (2003) 1579-1554
4. Xiao-Yuan Huang, Zi-Bin Dai: Design on FPGA Implementation of AES Algorithm chip. Microelectronics and computer, vol. 28(2005)62-68
5. Yeong-Kang Lai, Li-Chung Chang, Lien-Fei Chen, Chi-Chung Chou, Chun-Wei Chiu: A Novel Memoryless AES Cipher Architecture for Networking Applications. International Symposium on Circuit and systems, Vancouver, Canada (2004)23-26

Fault Diagnosis of Complicated Machinery System Based on Genetic Algorithm and Fuzzy RBF Neural Network*

Guang Yang[1,2], Xiaoping Wu[1], Yexin Song[3], and Yinchun Chen[4]

[1] Department of Information Security, Naval University of Engineering, 430033 Wuhan, China
sunyoung0783@sina.com
[2] Department of Watercraft Engineering, Zhenjiang Watercraft College, 212003 Zhenjiang, China
sunyoung0783@163.com
[3] Department of Applied Mathematics, Naval University of Engineering, 430033 Wuhan, China
yxsong@21cn.com
[4] Department of Control Science & Technology, Huazhong University of Science & Technology, 430074 Wuhan, China
chyc@21cn.com

Abstract. Compared with traditional Back Propagation (BP) neural network, the advantages of fuzzy neural network in fault diagnosis are analyzed. A new diagnosis method based on genetic algorithm (GA) and fuzzy Radial Basis Function (RBF) neural network is presented for complicated machinery system. Fuzzy membership functions are obtained by using RBF neural network, and then genetic algorithm is applied to train fuzzy RBF neural network. The trained fuzzy RBF neural network is used for fault diagnosis of ship main power system. Diagnostic results indicate that the method is of good generalization performance and expansibility. It can significantly improve the diagnostic precision.

1 Introduction

Many achievements have been obtained in applying BP network to fault diagnosis [1]. However, many problems remain to be solved in choosing network structure, training network, improving network generalization performance and so on, which limit its wide application in engineering practice. Especially, the number of training samples and local minima greatly affect network generalization performance, so the reliability of diagnostic results is significantly decreased. Fuzzy Neural Network (FNN) is a combination of fuzzy system and neural network. It enjoys not only the advantage of neural network's numerical computation, but also the characteristic of fuzzy system handling expert knowledge and diagnosing multiple faults simultaneously. So it is

* This work is partially supported by CNSF Grant #70471031.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 908 – 917, 2006.

suitable for fault diagnosis of complicated system. Despite their differences in topology structure, the most difference between FNN and traditional BP network is that the former has definite physical meanings on handling inputs and outputs [2]. For a fault diagnosis system, the fuzzification of inputs is the contribution degree of the measured physical variables to fault causes, while network outputs are the measurement of the input diagnostic vector belonging to a certain fault after the adaptation degrees of all diagnostic regulation are integrated by using the centroid method. Its input-output mapping is the mapping between the interval of diagnostic vector and the degree of belonging to a certain fault. However, in traditional BP network, its input-output mapping is the mapping between a fault eigenvector and Boolean value of whether a fault occurs or not. Each layer of its network has no physical meanings and only obtains numerical value relation through the training of samples. Therefore, FNN has better generalization performance than traditional BP network and the diagnostic result of FNN is more reliable than that of BP network.

Fuzzy BP neural network has found many applications in fault diagnosis [3, 4]. However, fuzzy BP neural network has its unique characteristics, so does fault diagnosis in different applications. Many problems remain to be solved, such as the method of determining appropriate fuzzy membership functions for input vectors and of improving learning rate and convergence performance of fuzzy BP neural network. In order to solve these two problems, based on the analysis of FNN, fuzzy membership functions can be obtained through RBF neural network and then genetic algorithm (GA) is used to train RBF neural network. Then the fuzzy RBF neural network is used for fault diagnosis of ship main power system. Diagnostic results indicate that the model is feasible and valid. With good generalization performance, the network has significantly improved the diagnostic precision of complicated system.

2 Diagnosis Model

2.1 Ascertaining Fuzzy Membership Functions

Usually, a fuzzy membership function is chosen based on experience, resulting in many shortcomings. Firstly, it is difficult to ascertain whether the type of the chosen membership function can reasonably fuzzify the input physical variables. In general, a function is chosen subjectively for the fuzzification of inputs, which may result in great difference between fuzzification results and practical conditions. In addition, even if the type of the chosen membership function can correctly describe the fuzzification rules of inputs, the function parameters ascertained after training are only consistent with network outputs in numerical value. They do not necessarily represent the fuzzification regulations of inputs.

RBF network is a common three-layer feedback network, with its structure shown in Fig. 1. The input layer nodes are responsible for transmitting input data to hidden layer nodes. Hidden layer nodes, i.e., RBF nodes, are composed of radiation functions like Gaussian kernel function, while simple linear functions are adopted in output nodes. The function (kernel function) in hidden layer nodes will locally respond towards input signals. I.e. when the input signals are close to the center of the function

range, the hidden layer nodes will produce bigger outputs; otherwise smaller ones. Gaussian function is the most used in general, which can be written as follows:

$$u_j = \exp[-\frac{(X-C_j)^T(X-C_j)}{2\sigma_j^2}] \quad j=1,2,\cdots,h \tag{1}$$

Where: u_j is the output of j th hidden layer node and $u_j \in [0,1]$; $X=(x_1,x_2,\cdots,x_n)^T$ is the input sample; C_j is the central value of Gauss kernel function; σ_j is the standardization constant; h is the number of hidden layer RBF nodes.

The output of RBF network is a linear combination of the output of hidden layer nodes, i.e.

$$y_i = \sum_{j=1}^{h} w_{ij} \cdot u_j = W_i^T \cdot U \ ,(i=1,2,\cdots,m) \tag{2}$$

Where $W_i=(w_{i1},w_{i2},\cdots,w_{ih})^T$ $,U=(u_1,u_2,\cdots,u_h,)^T$

The learning process of RBF network can be divided into two stages. Firstly C_j and σ_j of each hidden layer node are determined according to all input samples. After the parameters of hidden layer are determined, the weight vector W_i of output layer can be attained using least square method according to the samples.

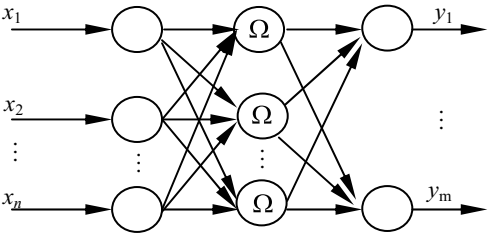


Fig. 1. The structure of RBF network

In the fault diagnosis domain, RBF neural network is one of the most rapidly developed forward networks in these years. Compared with other networks, it has the advantage of rapid convergence rate of learning algorithms (it is 10^3~10^4 times as rapid as the current widely used BP algorithm). It has been proved in theory that given any arbitrary $\varepsilon > 0$ and a continuous function vector h, of which each element satisfies the existence of $\int_{[0,1]^n} |h_i(x)|^2 dx$ and $h:[0,1]^n \subset R^n \to R^m$, then there exists a three-layer RBF neural network with single hidden layer that can approach to the function h with the approaching error less than ε. So RBF network can approach to an arbitrary continuous function within any given precision [5, 6]. Research indicates Gaussian Basis Function has the best approximate performance [7]. Whether training samples or not training samples, the approaching precision of RBF network is higher than that of BP network [8]. In the paper, fuzzy membership functions are got using

RBF neural network instead of by artificial selection, consequently selection of membership functions is optimized. Therefore the input fuzzication of fuzzy neural network is realized.

2.2 RBF Network Learning Algorithm Based on GA

Present RBF network learning algorithm based on gradient algorithm has shortages as follows:

Firstly, from the viewpoint of mathematics, it is a nonlinear gradient optimization problem. So local minima exist inevitably.

Secondly, limited by gradient algorithm, learning convergence rate is too slow with too much iterative time.

Genetic algorithm (GA) is introduced to RBF neural network learning process, thus the network not only keeps its original character, but also increases its convergence rate so as to reach global optimal solution or near optimal solution.

GA is an optimization algorithm inspired by biological evolution [9], which adopts simple encoding technique to denote all kinds of complicated structure, supervises learning and determines the searching direction through simple genetic operation and natural selection of a group of encodings. Using the population mode to search, it searches multi-zone in solution space simultaneously, making it suitable for large-scale parallel computation. So it can quickly constringe to global optimal solution or near optimal solution without restriction of search space (such as differentiability, continuity, single apex etc). From the start of initialization population, selection strategy based on fitness proportion is adopted to select individuals in current population, and reproduction, crossover and mutation operation are randomly used to produce next generation population until the genetic stopping conditions is satisfied.

The method of genetic computation is as follows

Binary-coding and real-coding are commonly used in gene encoding. Because binary-coding is not as good as real-coding in computation speed and precision [10], the real-coding is adopted in the paper. The network's error function is used to compute fitness function and the individual which has bigger error is considered to have smaller fitness. It can be shown as following:

$$f_d = \frac{1}{1+E_d} \tag{3}$$

Where f_d is the d-th individual's fitness function, E_d is the network's error function and it can be denoted as

$$E_d = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}(c_j^i - c_j)^2 \tag{4}$$

Where m is the sample number of training set and n is the neuron number of output layer. c_j is the practical output of the j-th output layer cell and c_j^i is the ideal output of the j-th output layer cell.

Step 1. Population $P(0)=\{P_1,P_2,\cdots,P_D\}$ is randomly initialized in the interval $[a_p,b_p]$ of population's individual domain, where the population size D is even and the genetic generation $h=0$;

Step 2. Compute h-th generation criterion (error) function $E_d(h), d=1,2,\cdots,D$;

Step 3. Comparing $E_d(h)$ with the threshold value $E_{\min}$, if $E_d(h)<E_{\min}$, the training stops, otherwise compute individual fitness $f_d(h)=\dfrac{1}{1+E_d(h)}$. Then $P_d(h)$ is arranged according to $f_1(h)\geq f_2(h)\geq\cdots\geq f_D(h)$ over again and the selection probability $P_d(h)=\left|\alpha-\dfrac{2(\alpha-1)}{D+1}\right|/D$ is computed, where α is a constant and $1\leq\alpha\leq 2$;

Step 4. Repeat the following computation $\dfrac{D}{2}$ times;

a. Randomly select two parent solution $P_E(h)$ and $P_L(h)$ according to the selection probability;

b. Take a random number r in the interval (0,1), if $r<p_r$, then carry out reproduction operation; if $p_r<r<p_c$, carry out crossover operation:

$$P_E(h+1)=P_E(h)+r_c[P_L(h)-P_E(h)] \tag{5}$$

$$P_L(h+1)=P_L(h)+r_c[P_E(h)-P_L(h)] \tag{6}$$

Where p_r and p_c are the upper bound of reproduction and crossover probability respectively; r_c is a random number within the interval (0,1). In order to quicken convergence rate, when both $\|f_L(h)-f_E(h)\|$ and $\|P_L(h)-P_E(h)\|$ are small, $r_c=k\cdot r_c$ can be chosen (k is an integer). When $p_c<r<1$, mutation operation is carried out on the two parent solutions respectively. A random positive integer m is generated. If m is even, then:

$$P_{Em}(h+1)=P_{Em}(h)+r_c[b_p-P_{Em}(h)](1-r_v^{T_E(h)^\lambda}) \tag{7}$$

$$P_{Lm}(h+1)=P_{Lm}(h)+r_c[b_p-P_{Lm}(h)](1-r_v^{T_L(h)^\lambda}) \tag{8}$$

If m is odd, then:

$$P_{Em}(h+1)=P_{Em}(h)+r_c[a_p-P_{Em}(h)](1-r_v^{T_E(h)^\lambda}) \tag{9}$$

$$P_{Lm}(h+1)=P_{Lm}(h)+r_c[a_p-P_{Lm}(h)](1-r_v^{T_L(h)^\lambda}) \tag{10}$$

Where $T_E(h)=1-\frac{f_E(h)}{f_1(h)}$, $T_L(h)=1-\frac{f_L(h)}{f_1(h)}$, $\lambda=2\sim5$, r_v is a random number within interval (0,1).

Step 5. $2D$ individuals made up of $P_d(h)$ and $P_d(h+1)$ are ordered according to their fitness and the top D individuals are selected as the new $P_d(h+1)$;

Step 6. Turn to Step 3.

3 Diagnostic Example

An illustrative example is given for fault diagnosis of combustion subsystem with six basic fault phenomena, denoted as $F=\{F_1,F_2,F_3,F_4,F_5,F_6\}$, where F_1 denotes that the injector's nozzle enlarges; F_2 denotes that the injector's nozzle is jammed; F_3 denotes that the valve seat of injector leaks; F_4 denotes that fuel is sprayed too late; F_5 denotes that fuel is sprayed too early; F_6 denotes that the vent-pipe is jammed. Analysis of fault mechanism of combustion subsystem in ship main power system and practical examples of faults indicates that all the faults F_1~F_6 are related with the maximum explosion pressure p_z and the exhaust temperature T_r. Therefore, p_z and T_r can be chosen as diagnostic characteristic parameters.

3.1 Obtaining Input Membership Functions

Knowledge of fault diagnosis has to be represented as pattern of fuzzy rule before constructing fault diagnosis system based on FNN. Then FNN learning samples will be constructed according to fuzzy rule. Diagnostic characteristic parameters should be fuzzified for the fault diagnosis fuzzy rule. Diagnostic characteristic parameters are defined as five fuzzy sets of "too big/high", "big /high", "normal", "small/low" and "too small/low". Training samples are made up of normalized values of diagnostic parameters, and membership degree of the corresponding fuzzy sets. The normalized formula is defined as follows.

$$\tilde{Z}=\frac{Z-Z_{\min}}{Z_{\max}-Z_{\min}} \tag{11}$$

Where Z is the actual value of a diagnostic parameter; $Z_{\max}$ and $Z_{\min}$ are maximum and minimum of a diagnostic parameter respectively according to expert experience; $\tilde{Z}$ is the normalized value.

According to domain expert experience, the parameter ranges of the maximum explosion pressure p_z and the exhaust temperature T_r are 123×10^5Pa~137×10^5Pa and 320^0C~360^0C respectively. Each network structure is 2-5-5, i.e., two input nodes, five hidden layer nodes and five output nodes. In the training samples, the inputs are the normalized values of diagnostic parameters and the outputs are the corresponding membership degree of five elements of fuzzy set, i.e., "too big/high", "big/high", "normal", "small/low" and "too small/low". When network training finishes, the corresponding membership functions of the maximum explosion pressure and the exhaust temperature can be obtained.

3.2 Training of Fuzzy RBF Network

After the corresponding membership functions of the maximum explosion pressure and the exhaust temperature are obtained, the remaining task is to train the network with the numerical training sets corresponding to fault reasoning rules determined by expert experience. It is a very important and difficult task to attain the training sets for neural network fault diagnosis. In order to construct training sample sets, we should combine a great deal of expert knowledge about faults and valuable experience about fault diagnosis for main power combustion system. In this paper, combining much expert knowledge about machinery faults of power system and practical experience of ship main power combustion system, the main expert reasoning rules are constructed as are shown in Table 1, where "seldom", "accidentally", "sometimes" and "often" respectively denote the frequency of a fault occurring. GA is used to train fuzzy RBF network, and the training error curve is shown in Fig. 2. When the mean square error reaches 0.001, network training finishes. At the same time, in order to illuminate the advantage of network training based on GA, the training error curve of fuzzy RBF network based on gradient algorithm is shown in Fig. 3. Compared Fig. 2 with Fig. 3, it can be seen that convergence rate using GA is greatly improved and the network is not easy to converge on local minima.

Table 1. Fault diagnosis basic rules in combustion system

Fuzzy symptom parameters (condition)		Faults (conclusion)					
P_z	T_r	F_1	F_2	F_3	F_4	F_5	F_6
too big	too high	sometimes	seldom	seldom	seldom	seldom	sometimes
normal	too high	seldom	seldom	sometimes	accidentally	seldom	often
too small	too high	often	seldom	often	often	seldom	seldom
too big	normal	seldom	seldom	seldom	seldom	sometimes	accidentally
normal	normal	seldom	seldom	seldom	seldom	seldom	seldom
too small	normal	seldom	seldom	seldom	sometimes	seldom	seldom
too big	too low	seldom	seldom	seldom	seldom	often	seldom
normal	too low	seldom	accidentally	seldom	seldom	seldom	seldom
too small	too low	seldom	often	seldom	seldom	seldom	seldom

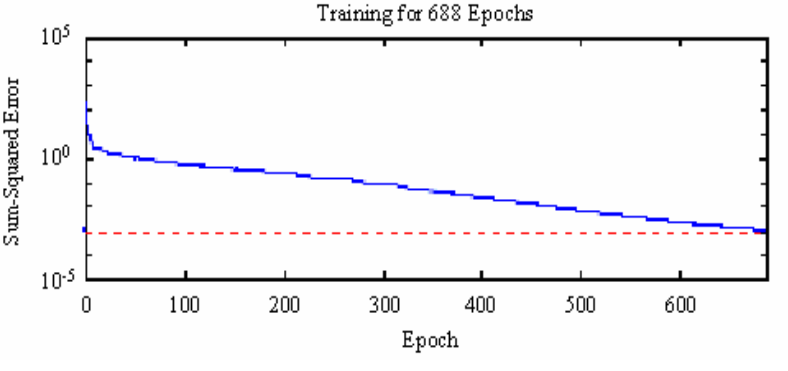


Fig. 2. Training error curve of fuzzy RBF network based on GA

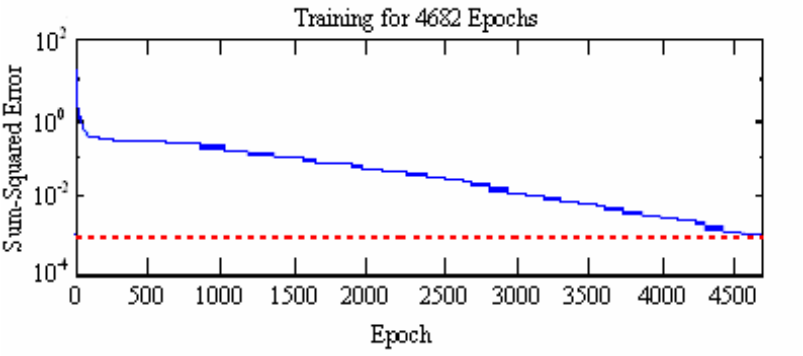


Fig. 3. Training error curve of fuzzy RBF network based on gradient algorithm

Table 2. Comparison of preciseness rate of diagnostic results

Diagnostic method	Diagnostic preciseness rate of single fault occurring	Diagnostic preciseness rate of two or more faults occurring
Fuzzy RBF NN based on GA	98%	90%
Fuzzy BP NN based on BP	94%	80%

Table 3. Comparison of partial diagnostic results

Input (symptom parameters)			Diagnostic	Output (fault credibility)					
Parameters	P_z	T_r	method	F_1	F_2	F_3	F_4	F_5	F_6
Same samples	too big (0.90)	too low (0.10)	Fuzzy RBF NN	0.0507	0.0498	0.0536	0.0467	0.9481	0.0587
			Fuzzy BP NN	0.0512	0.0531	0.0608	0.0314	0.9025	0.0547
	too small (0.10)	too low (0.10)	Fuzzy RBF NN	0.0532	0.9538	0.0516	0.0564	0.0602	0.0497
			Fuzzy BP NN	0.0543	0.8976	0.0598	0.0604	0.0717	0.0513
Similar samples	0.80	0.80	Fuzzy RBF NN	0.5129	0.1557	0.1215	0.0436	0.0987	0.5768
			Fuzzy BP NN	0.4708	0.3497	0.2013	0.0615	0.1834	0.6503
	0.25	0.80	Fuzzy RBF NN	0.6224	0.0573	0.7130	0.6893	0.0357	0.1342
			Fuzzy BP NN	0.4126	0.0617	0.7632	0.6514	0.0408	0.2546
Different samples	0.65	0.35	Fuzzy RBF NN	0.1375	0.0988	0.2056	0.0875	0.6835	0.0137
			Fuzzy BP NN	0.4327	0.1384	0.3789	0.2674	0.5072	0.0983
	0.40	0.40	Fuzzy RBF NN	0.1578	0.5921	0.1029	0.0547	0.1892	0.1349
			Fuzzy BP NN	0.1862	0.1249	0.1591	0.0632	0.0918	0.1014

3.3 Diagnostic Results

According to the above training results, the fuzzy RBF neural network fault diagnosis system is simulated with 100 groups of practical data. The diagnostic performance is compared between GA-based fuzzy RBF neural network and BP-based fuzzy neural

network. The comparison of preciseness rate is shown in Table 2. Some output results are shown in Table 3. From Table 2, it can be seen that the diagnostic preciseness rate of GA-based method is much higher than that of BP-based method. It can also be found from the simulation that, with the same samples and similar samples, good fault diagnosis result can be obtained using both methods, but the former is more precise than the latter. However, except that fuzzy BP neural network is easy to converge on local minima when the number of middle layer nodes is not suitable, it has bad results for the diagnostic aspect of non-fault samples. Miss or wrong judgments are often given. The diagnosis for different samples' fault F_2 (the injector's nozzle is jammed) in Table 3 can clearly explain those. The fault credibility degree of output F_2 is 0.5921, denoting that the fault is in a standard fault stage. However, ambiguous results are given by fuzzy BP neural network method and the values of its output F_1, F_2 and F_3 are almost of the same. If the maximum is chosen as the standard of fault, then fault F_1 (the injector's nozzle enlarges) is diagnosed by fuzzy BP neural network. Obviously this is a wrong judgment.

4 Conclusions

On the analysis of problems in the application of fault diagnosis with fuzzy neural network and traditional neural network based on BP algorithm, a new method using RBF network to get membership functions is presented. The training of fuzzy neural network is improved by using genetic algorithm, which makes it easier for the extension and update of diagnostic network's expert knowledge. The network has good generalization performance. Simulation results indicate that with few or no wrong judgment, the method has better diagnostic accuracy than fuzzy BP neural network. It has rapid convergence rate and seldom converges on local minima during training.

Fault diagnosis model based on GA and RBF neural network can also be used in other complicated systems, which gives ship main power system a beneficial reference for intelligent diagnosis and paves the way for the fault diagnosis expert system based on fuzzy neural network. However, the method also has its shortages which should be solved in the future. I.e., network training has the randomicity when using GA, whose convergence rate changes greatly according to different initialization values. This shortage can be solved only through improving GA continuously. Meanwhile, how to obtain the right fault reasoning rules to form the training sets for FNN fault diagnosis is still a very vital and difficult task.

References

1. Hu, K. C., Kong, F. R., Li, C. Q.: Hidden Layer's Compression Algorithm for Neural Network-Based Diagnosis Models. In: Journal of Vibration Engineering. Vol. 10, No. 3. Jiangsu (1997) 356-361
2. Zhang, J. X., Zhong, Q. H., Dai, Y. P.: Fault Diagnosis of Rotating Machinery Using RBF and Fuzzy Neural Network. In: Journal of System Simulation. Vol. 16, No.3. Beijing (2004) 560-563
3. Wu, J. H., Yang, X. L.: Fuzzy BP Neural Network and Its Applications in Fault Diagnosis. In: Systems Engineering and Electronics. Vol. 23, No.10. Beijing (2001) 560-563

4. Yao, H. X., Zhao, L. D., Sheng, Z. H.: Application of Multi-grade Fuzzy Neural Networks in Fault Diagnosis of Large Machinery. In: Journal of Southeast University (Natural Science Edition). Vol. 31, No. 2. Nanjing (2001) 59-63
5. Wang, L. X., Mendel, J. M.: Fuzzy Basis Functions, Universal Approximation, and Orthogonal Least-squares Learning. In: IEEE Trans. on Neural Networks, Vol. 3., (1992) 807-814.
6. Hornik, K., Stinchcombe, M., White, H.: Multilayer Feed-forward Network are Universal Approximators. In: Neural Network. Vol. 2, No.6. (1989) 359-366
7. Poggio, T., Girosi, F.: Networks for Approximation and Learning. In: Proc. IEEE, Vol. 78. (1990) 1479-1481
8. A Comparative Study on BP Network and RBF Network in Function Approximation. In: Bulletin of Science and Technology. Vol. 21, No.2. (2005) 193-197
9. Bruse, A.: Genetic Evolution of Radial Basis Function Coverage Using Orthogonal Niches In: IEEE Trans. on Neural Networks. Vol. 7, No.6. (1996) 1528-1996
10. Hu, H. Y., Zhu, S. W., Zhang, D. B., et al: Oblique Decision Tree Construction with Decimal-coded Genetic Algorithm. In: Computer Science. Vol. 28, No.2 (2001) 108-110

An Infrared and Neuro-Fuzzy-Based Approach for Identification and Classification of Road Markings

G.N. Marichal, E.J. González, L. Acosta, J. Toledo, M. Sigut, and J. Felipe

Dep. Ingeniería de Sistemas y Automática y Arquitectura y Tecnología de Computadores.
Universidad de La Laguna, CP 38207, Tenerife, Spain
{nicomar, ejgonzal, lacosta, jttoledo, marsigut}@ull.es

Abstract. A low-cost infrared and Neuro-Fuzzy-based approach for identification and classification of road markings is presented in this paper. The main task of the designed prototype, implemented at the University of La Laguna, is to collect information by a set of infrared sensors and to process it in an intelligent mode. As an example, it should inform the driver about the different signs and elements it is driving on, complementing other well-known and widely used security devices. For the identification and classification stages, a Neuro-Fuzzy approach has been chosen; given it is a good option in order to get fuzzy rules and memberships functions in an automatic way.

1 Introduction

Sensibility to traffic concerns such as intelligent transport, energy consumption, etc. has experimented an increase last decade [1],[2]. In particular, several researchers have paid special attention to safety, as less traffic accidents involve less traffic jams, fewer injuries and less economic loss. In this context, traffic sign detection and classification (both as driver assistance and as a support for intelligent vehicles) has become a target for many research groups. Regarding to classification, it is usually preceded by a normalization procedure. The resulting candidate is entered into a classifier tool, such as nearest neighbour, neural networks, genetic algorithms, template matching, etc [3],[4], [5]. It is necessary to carry out quickly the recognition process in order to help the human driver, making driving easier and safer.

Most of the proposed implementations make use of vision, integrating one or two cameras in the vehicle [6]. However, implemented methods are affected by the large degree of freedom existing in a traffic scenario. Apart from vision-based approaches, there are other devices that are used for that purpose. In particular, the authors are interested in infrared technology. Although this technology has been applied to road traffic field (for example in the determination of the flow of vehicles in a specific road), in this paper, the authors are interested in its application to the design of vehicle security devices.

The remainder of the paper is as follows: In Section 2, a low-cost prototype for the detection and identification of traffic signs or elements painted on the road is briefly described. In Section 3, sensorial information processing (including data pre-processing and

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 918 – 927, 2006.

Neuro-Fuzzy approach) in order to classify the traffic signs is detailed. The application of the Neuro-fuzzy approach to the classification of road markings and obtained results are discussed in Section 4. Finally, conclusions are reported in Section 5.

2 Infrared System

The main task of the designed prototype, implemented at the University of La Laguna, is to collect information by a set of infrared sensors and to process it in an intelligent mode. As an example, it should inform the driver about the different signs and elements it is driving on, complementing other well-known and widely used security devices. Thus, this prototype could advise about an undesired movement to other lane, that is, when the corresponding direction indicator has not been activated by the driver. Nevertheless, its application is not limited to that case, but it is able to identify different signs on the road (turn right, turn left, yield sign, etc.). An application of this feature in the field of Robotics would consist in the interaction of mobile robots in real environments such as car parks.

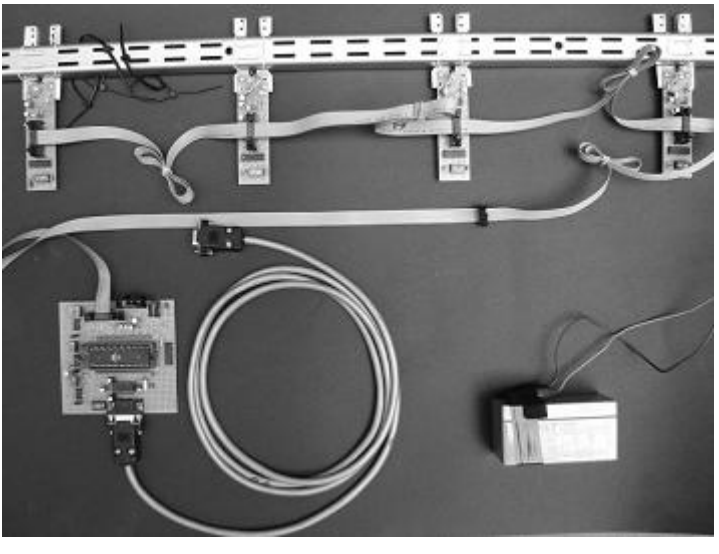


Fig. 1. Structure of the prototype

The prototype is a really low-cost device as its price is about $70. This is an interesting feature when comparing this prototype with vision-based systems, in spite of the decreasing prices of the video cameras. Moreover, the space it takes up is not significant. The device consists of a set of infrared emitters and sensors, able to detect the reflection on the road. In other words, the prototype differentiates between the painted zones and the rest. Each sensor is able to cover a circular area indicating if it is a painted zone or not. Empirically, for an average height of 15 cm, the sensor covers a circular area of R=10 cm. The prototype consists of three parts:

- Emitter/receiver pairs. They are located in a bar, across the vehicle. Each emitter/receiver pair is integrated in a circuit board together with a switch circuit that serves as identification element. The infrared sensor has been encapsulated into an aluminum-made wrapping with several holes. These small holes, made in a specific direction, allow that only a reflected infrared wave (generated by the infrared emitter and reflected on a painted surface) could activate the sensor device. In this prototype, the wrapping is made of two aluminum pieces. These pieces are jointed with black silicone, keeping the opaqueness.
- Control board. It includes an oscillator circuit for the emitter circuit and a microcontroller that governs the successive data obtained by the sensors and sends that information to a portable computer.
- Portable computer. It processes the data received from the microcontroller, representing it as an image that represents what the sensors have registered from the road. As the reception is binary (there is reflection or not), that representation can be formed of black and white blocks.

The system will make a complete cycle, sending the read signal to each transmitter and reading the output from the corresponding sensor. This cycle is repeated while the vehicle is circulating, obtaining a line-to-line image from the road. After receiving the data from the microcontroller, the computer looks for the traffic sign that fits better with those data. For that purpose, it uses a system based on Neuro-Fuzzy theory, described in the following section.

3 Sensorial Information Processing

Following the process depicted in the previous section, the infrared system is able to take 5 data each time, given 5 infrared sensors are located in the infrared system. Once the vehicle is moving along the sign more data are read, our proposal is to process these data in order to obtain which sign is painted on the street. At the first stage of the work, a Fuzzy System was considered, however, the difficulty in finding out the specific fuzzy rules and memberships function makes this approach cumbersome in this case. On the contrary, a Neuro-Fuzzy approach has been chosen; given it is a good option in order to get these fuzzy rules and memberships functions in an automatic way. In the case treated in this paper a Neuro-Fuzzy system based on the Sugeno Approach has been used [7]. As it is well known the Sugeno approach is more convenient in terms of computational charge than the Mamdani approach [8].

3.1 Data Pre-processing

First problem in the application of the Neuro-Fuzzy approach is the number of data generated by the infrared system. In a typical case a matrix of 150 x 5 data could be generated for a typical sign, once a segmentation process has been carried out. This segmentation process is very simple, removing those rows without any valid detection and isolating the possible candidates. Due to this simplicity, the process is immediate. This is a critical feature, since the traffic sign only can be identified once the system - integrated in the vehicle - has read it.

As can be noted, the resulting matrix is usually too large to be processed properly. Hence, the first step is to simplify this matrix. In the case presented in this paper, it has been considered to divide this matrix in 5 x 5 windows. That is, each window corresponds to one region of the candidate to analyze, such that a number corresponding to the quotient between the number of valid detections and the total number of ones and zeros is associated. In this way, the output of the infrared system is converted into a matrix of 5 x 5. Therefore, the number of inputs of the Neuro-Fuzzy is 25, associating each input with each window.

On the other hand, only one output is necessary, where a specific number is linked to each traffic sign.

3.2 Structure of the Neuro-Fuzzy System

The used Neuro-Fuzzy System could be seen as a three-layer Network [9] [10]. The nodes of the network are cascaded together in layers.

The first layer or input layer comprises several nodes, each one consisting of a Radial Basis neuron. As it can be seen, this particular choice of the nodes at the input layer is linked to the fact that Gaussian membership functions are considered for the input variables. Note as the inputs variables in this case, are the quotients, which were referred in the previous sub-section.

In the hidden layer a typical AND operation (obtaining the minimum between the operands) is applied to the membership functions of the inputs. Because of that, this layer could be known as rule layer.

Finally, the output layer could be considered as a linear neuron layer, where the weight connections between the hidden layer and the output layer are the estimated values of the outputs. Hence, it could be said that the output layer carries out the defuzzification process, providing the outputs of the Neuro-Fuzzy System.

To sum up, the structure of the Neuro-Fuzzy system could be seen as a typical Radial Basis Network, where an additional layer has been inserted between the Radial Basis layer (the input layer) and the linear layer (the output layer). The neurons of this additional layer calculate the degrees of membership corresponding to the different rules, that is, they apply the fuzzy operator AND (min operator), being the total number of nodes, the total number of Fuzzy rules. Once, these calculations have been carried out, the output layer applies a defuzzification process in order to obtain numeric values for the outputs of the system. Note as this structure allows to identify the different rules against a pure Neural Network approach.

In the Neuro-Fuzzy System, there are some parameters which determine the relation between the inputs and outputs. In this case, the behaviour of the Neuro-Fuzzy system depends on the value of the following parameters: the membership function centres, the widths of the membership functions and the estimated output values. In order to determine these parameters a learning algorithm has to be designed.

3.3 Learning Algorithm

The learning algorithm can be divided into three phases. First phase consists of obtaining convenient initial values for the membership function centres and estimated output values. A Kohonen's Self-organising feature map algorithm [10] has been applied. It looks for these initial values, using the pairs 5x5 windows and the corre-

sponding traffic signs as training patterns. Second phase is focused on optimising the number of nodes at the hidden layer, that is, the number of Fuzzy rules. Whereas the last phase is focused on adjusting the parameters of the Neuro-Fuzzy system, and improving the given ones in the previous phases. Taking into account the similarities between the typical Radial Basis Network and the Neuro-Fuzzy system, the least mean squared learning algorithm has been used as usual in a typical Radial Basis Network [10].

4 Application

In order to test the methods, the infrared system was installed under a vehicle as it is shown in Figure 2.

Fig. 2. Photographs of the used vehicle for testing the system

In the first stage of this work, the set of traffic signs shown in Figure 3 (a) has been pre-processed as it was pointed out in section 3.1. In Figure 3, it is also shown the corresponding pre-processed image of each sign in the ideal case, when all parts of the sign are captured by the infrared system in a perfect way. Furthermore, these images have been chosen as the inverted traffic sign images, given the data are collected from the beginning of the traffic sign up to the end, when the vehicle is passing above it. Several captures of the same traffic sign were done, obtaining in general different results of the pre-processed images shown in Figure 3 (b). This set of traffic signs has been divided into two sets: the training set and the test set. After this step, several trials have been carried out, using the training set and the Neuro-Fuzzy approach. The outputs of the Neuro-Fuzzy have been chosen as it is indicated in Table 1. Note as different values have been associated for each traffic sign and all traffic signs have its own output value. However, only two outputs values were used at the beginning. One output value for one of the traffic signs and another for the rest of them. That is, one Neuro-Fuzzy system was chosen for each traffic sign to be classified. This approach made the training process more difficult. In general unstability in the training process and worse results were achieved. Because of that, one Neuro-Fuzzy system with one output value for each type of traffic sign was preferred. Sum squared errors of $3.5\ 10^{-4}$ have been obtained, after 1000 epochs.

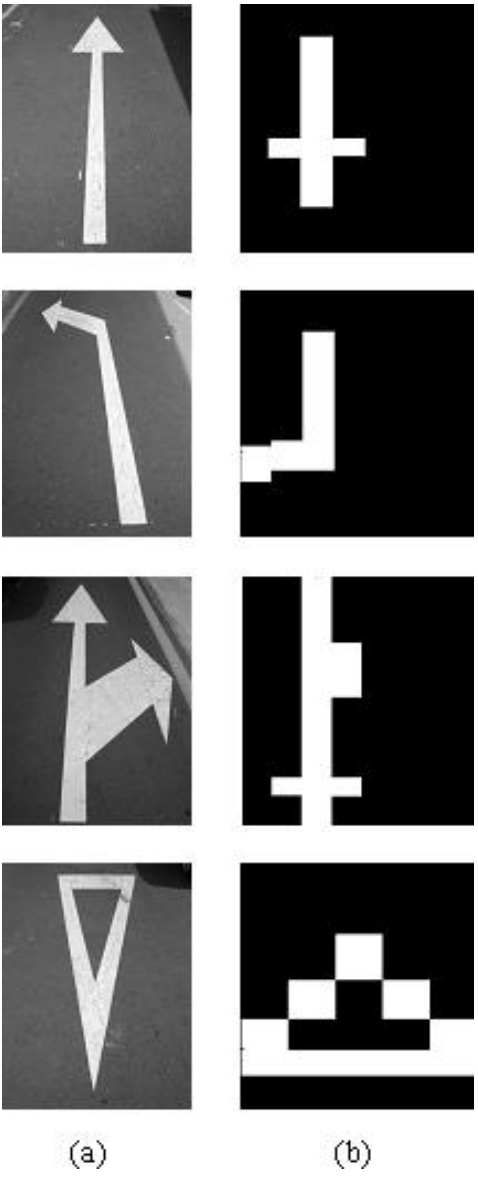


Fig. 3. (a) Photographs of the traffic signs used in the experiments and (b) the obtained pre-processed images in the ideal case, when all infrared sensors captures without defect the traffic sign

Once the training phase has been finished, a way of determining the generality degree of the Neuro-Fuzzy system is to use the test set. The outputs for the trained Neuro-Fuzzy system corresponding to the test set are shown in Table 2. As it can be seen the outputs are around the output reference values for each traffic sign. That is, a Neuro-Fuzzy system with fixed parameters and good generality capabilities has been achieved.

Table 1. Reference output values of the traffic signs

Traffic sign	Reference output
Arrow	1
Right Arrow	3
Yield sign	5
Forward-right Arrow	7

Table 2. Outputs for the test patterns

Test sign pattern	Output value
Arrow test pattern 1	0.9973
Right test pattern 1	2.9800
Right test pattern 2	2.9763
Yield sign test pattern 1	5.3931
Yield sign test pattern 2	4.7881
Forward-right Arrow test pattern 1	7.0572

Note as the resultant Neuro-Fuzzy system could be considered as a typical Sugeno type Fuzzy System. Hence, a set of Fuzzy rules could be extracted, being one of the main advantages of this approach versus others, as Neural Netwok paradigm [9], [11]. However, one of the disadvantages of the Neuro-Fuzzy system is the high number of rules obtained in most times. In the case presented in this paper, a Fuzzy System built up 25 rules has been achieved after training process. Note as the training process includes a rule minimization phase.

It would be interesting to extract this rule base for the resultant Neuro-Fuzzy system, in order to know what these rules look like, which of them are important in the whole classification task and which of them are more important to classify a specific traffic sign.

On the other hand, this knowledge would allow an expert to add new rules or to modify some of them in order to get a system more adjusted to the requirements of the specific environment. Different methods could be devised in order to get this target. In this paper, different parameters of the Neuro-Fuzzy system have been considered for determining the influence of the rules to a specific situation.

In the following, the trained Neuro-Fuzzy system will be considered. First step in order to get the influence degree of the rules could be to study the estimated values, pointed out in the following as $sv_{i,j}$.

In Table 3 the rules associated with each traffic sign, according to the output reference value, are presented. Furthermore, an additional column is shown, where values inside the range [-1 0] are shown. Evidently, it is not clear which traffic signs are associated with. Moreover, although all the rules corresponding to a column could contribute to classify that specific traffic sign, it is no secure all those rules contribute or at least they contribute at the same level. Hence, the second step in the process is to confirm with rules are important for each specific case. The chosen procedure in this paper has been to use the γ_j parameters (outputs of the nodes at the hidden layer) ,

corresponding to the test set. In the following, the γ_j parameter is rename as γ, for the sake of simplicity. It also will be considered as a percentage of rule influence, given γ_j is inside the range [0 1]. It could be seen that there are a set of rules with a γ parameter below 0.1 %. In fact, it could be seen that there is a sub-set of rules with γ less than 0.1 %, which are common to all patterns. This result tell us in a first approximation, that only 16 rules are important in the classification process from the 25 rules corresponding to the whole set. This is a first simplification of the number of rules. It seems reasonable to take only 16 rules, without taking into account the rules with a γ less than 0.1 %. Furthermore, it could be seen which of these 16 rules are more influential for the classification of a particular traffic sign. In Table 4, it is shown a summary of the most important rules at a first sight for each specific traffic sign, considering the results of evaluating the γ parameter for a wide set of patterns . In fact, Table 4 indicates with rules have a influence greater than 25% for classifying each traffic sign from the whole rules shown in Table 3. Additionally, Table 4 shows that the rules whose estimated values (sv) are inside the range [-1 0] also play an important role. Although their estimated values are far away from the output reference values. Table 4 could be seen as a second simplification of the number of rules. However, this simplification could be a oversimplication, that is, Table 4 could have a small number of rules to make a good classification. Because of that, a satisfactory set of Fuzzy rules could be considered the union set between rules in table 3 and table 4.

Table 3. Rules for each traffic sign according estimated values

	Arrow	**Right Arrow**	**Yield sign**	**Forward-right Arrow**	**Other Rules**
Interval	[0 2)	[2 4)	[4 6)	[6 8]	[-1 0]
Ref. Value	1	3	5	7	
Rules	6, 7	8,9,10,11,12	13,14,15, 16,17,18	19,20,21,22,23, 24,25	1,2,3,4,5

Table 4. Most important rules for each traffic sign according to γ parameters of a wide set of patterns

Pattern	**Rules**
Arrow	1 2 3 4 5 6 7 9 10 11
Right Arrow	7 9 10 11 13
Yield Sign	13
Forward-Right Arrow	22 23 24

5 Conclusions and Future Work

In this paper, an infrared and Neuro-Fuzzy-based approach for the identification and classification of road markings has been presented. For this purpose, the authors have

designed and implemented a prototype based on infrared emitters/receivers. Each emitter/receiver pair covers a circular area of road and it is able to detect if there is a painted road marking or not. The prototype should inform the driver about the different signs and elements it is driving on, complementing other well-known and widely used security devices. It is important to remark that the prototype is really low-cost, in spite of the decreasing prices of other possible devices such as video cameras.

After the data are collected, a Neuro-Fuzzy approach has been chosen for classification purposes; given this approach is a good option in order to get these fuzzy rules and memberships functions in an automatic way. In the case treated in this paper a Neuro-Fuzzy system based on the Sugeno Approach has been used.

In order to test the methods, the infrared system was installed under a vehicle. The set of collected data has been divided into two subsets: the training set and the test set, giving one output value for each type of traffic sign. With this premise, the Neuro-Fuzzy system is able to identify properly the elements of the test set. That is, a Neuro-Fuzzy system with fixed parameters and good generality capabilities has been achieved.

Apart from the related identification features, a set of Fuzzy rules has been extracted. The proposed system could be considered as a typical Sugeno type Fuzzy System. As an application, a study about a possible simplification of the Fuzzy rules set has been carried.

Future work includes a study about what the obtained rules are exactly doing, in order to an expert could modify these rules or design new rules more appropriate for a specific scenario.

Acknowledgement

This work has been financed by the Spanish government, project #DPI2004-06626.

References

1. Bertozzi, M., Broggi, A., Cellario, M., Fascioli, A., Lombardi, P., Porta, M.: Artificial Vision in Road Vehicles, Proceedings of the IEEE - Special issue on Technology and Tools: Visual Perception, (2002)
2. Gupte, S., Masoud, O., Papanikolopoulos, N.P.: Detection and classification of vehicles, IEEE Trans. on Intelligent Transportation Systems, Vol. 3, no. 1, (2002) 37-47
3. Tan, Y.P., Yap, K.H., Wang, L.P. (Eds.): Intelligent Multimedia Processing with Soft Computing. Springer, Berlin, Heidelberg, New York (2004).
4. Halgamuge, S., Wang, L.P. (eds.), Classification and Clustering for Knowledge Discovery, Springer, Berlin (2005).
5. Gavrila, D.M.: Traffic sign recognition revisited, Proc. 21st DAGM Symposium fur Musterekennung, Springer Verlag (1999) 86–93
6. de la Escalera, A., Armingol, J.M., Pastor, J.M., Rodriguez, F.J.: Visual sign information extraction and identification by deformable models for intelligent vehicles. IEEE Transactions on Intelligent Transportation Systems, Vol. 5 (2) (2004) 57- 68

7. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its applications to modelling and control. IEEE Trans. On Systems, Man and Cybernetics, Vol. 15 (1985) 116-132
8. Mamdani, E.H., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller, International Journal of Man-Machine Studies, Vol. 7 (1) (1975), 1-13.
9. Kosko B.: Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence. Prentice-Hall (1992)
10. Hush, D.R., Horne, B.G.: Progress in supervised Neural Networks. IEEE Signal Processing Magazine (1993) 8-34
11. Karvonen, J.A.: Baltic Sea ice SAR segmentation and classification using modified pulse-coupled neural networks. IEEE Trans. Geoscience and Remote Sensing, 42 (2004) 1566-1574

Unique State and Automatical Action Abstracting Based on Logical MDPs with Negation

Song Zhiwei and Chen Xiaoping

Computer Science Department, University of Science and Technology of China
96 Jinzhai Road, Hefei, Anhui 230026, China
songzw@ustc.edu.cn, xpchen@ustc.edu.cn

Abstract. In this paper we introduce negation into Logical Markov Decision Processes, which is a model of Relational Reinforcement Learning. In the new model nLMDP the abstract state space can be constructed in a simple way, so that a good property of complementarity holds. Prototype action is also introduced into the model. A distinct feature of the model is that applicable abstract actions can be obtained automatically with valid substitutions. Given a complementary abstract state space and a set of prototype actions, a model-free Θ-learing method is implemented for evaluating the state-action-substitution value funcion.

1 Introduction

It is widely agreed that the intelligent agent should have the ability of learning. *Reinforcement Learning* (RL) permits the agent to learn policy by interacting with environment without supervisors. Many algorithms of RL and the propositional representations of states have been studied extensively over the past years [1,2]. However, their usefulness in complex domains is limited because of their inability to incorporate relational information about the environment, especially in the domain with objects of various kinds. In order to solve this problem, *Relational Reinforcement Learning* (RRL) was proposed based on the relational representation [3]. It concerns use of RL in complex domains with the states and actions in relational form. Recently, a number of RRL algorithms have been developed [4,5,6,7,8] with several preliminary models proposed [9,10,11,12,13], but the problem of RRL is still not well-understood and the theory is not sufficient.

In this paper, a *Logical Markov Decision Process with Negation* (nLMDP) is proposed on the basis of *Logical Markov Decision Processes* (LOMDPs) [10,14]. The main differences between LOMDPs and our nLMDPs are the negation introduced in nLMDPs, the complementarity of abstract state space of nLMDPs, and the definition of the action space. The *negation* of logical languages is first introduced in abstract states of nLMDPs. Then the abstract state space can be constructed in a simple way and is complementary, that is each ground state can be represented by only one abstract state. In LOMDPs, a gound state can be represented by two or more abstract states, and using a manual pre-given total

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 928–937, 2006.

order of them to select the correct one. The action used in nLMDPs is also different from that of LOMDPs. It is a super abstraction over general abstract actions, called *Prototype Action*. Thus, applicable abstract actions of an abstract state can be gained automatically with valid substitutions. Then a value function of state-action-substituion is presented, denoted as Θ. Based on these, an algorithm of $\Theta(\lambda)$-*leaning* is implemented to evaluate the values of the substitutions.

This paper is organized as follows. Section 2 will briefly introduce some logical and MDP preliminaries. After proposing the nLMDP in section 3, the $\Theta(\lambda)$-learning is presented in section 4. Section 5 will give some experimental results of our method. Some related work are discussed in section 6, and the papers ends with conclusions and directions for further research.

2 Preliminaries

In this section, we briefly recall the terminology of logical languages and MDPs. For more details, see [15] and [2].

2.1 Logic

An *alphabet* Σ is a set of relation symbols, $\mathcal{P}$, and a set of *constants*, $\mathcal{C}$. If the arity $m(m \geq 0)$ of a relation symbols $\mathtt{p} \in \mathcal{P}$ is 0 then $\mathtt{p}$ is called a proposition. A *term* is a variable $\mathtt{X}$, or a constant $\mathtt{c}$. An *atom* $\mathtt{p}(\mathtt{t}_1, \cdots, \mathtt{t}_m)$ is a relation symbol $\mathtt{p}$ followed by a bracketed m-tuple of terms $\mathtt{t}_i$. A *conjunction* is a set of atoms. A *substitution* $\theta = \{\mathtt{X}_1/\mathtt{t}_1, \cdots, \mathtt{X}_n/\mathtt{t}_n\}$ is a set of assignments of terms $\mathtt{t}_i$ to variables $\mathtt{X}_i$. A conjunction A is said to be θ-subsumed by a conjunction B, denoted by $A \preceq_\theta B$, if there exists a substitution θ such that $B\theta \subseteq A$. A *fomula* is an atom or a conjunction. The variables of a term or a fomula A is denoted by $vars(A)$. The terms of a fomula A is denoted by $terms(A)$. A term, or a fomula is called *ground* when it contains no variables. To highlight A is ground, a bar is added, i.e. $\bar{A}$, in this paper. The *Herbrand base* of Σ, denoted as HB^Σ, is the set of all ground atoms which can be constructed with the predicate symbols and functor symbols of Σ. An *interpretation* is a subset of HB^Σ.

In logic programming, the logical *negation*, denoted by `not`, is often used. There is a little difference between it and the classical negation $\neg$ that `not` implies that there is a *closed world assumption* [16]. `not`A is true, iff A can not be inferred. For a conjunction A, `not`A is called *not-conjunction* in this paper. A not-conjunction is also a fomula. A set of several conjunctions and not-conjunctions is a fomula, too.

2.2 Markov Decision Processes

A *Markov Decision Process* is a tuple $M = \langle S, A, T, R \rangle$, where S is a set of states, A is a set of actions, $T : S \times A \times S \to [0, 1]$ is a *transition model*, and $R : S \times A \times S \to \mathbb{R}$ is a reward model with real values. A transition from state s to s' caused by action a occurs with a reward $R(s, a, s')$ and probability $T(s, a, s')$

which holds $\sum_{s'} T(s,a,s') = 1$. A *policy* $\pi : S \times A \to [0,1]$ is a mapping from states and actions to the probability $\pi(s,a)$ of taking action a when in state s.

Given an MDP $M = \langle S, A, T, R \rangle$, a policy π, and a *discount factor* $\gamma \in [0,1]$, the *state value function* $V^\pi : S \to \mathbb{R}$ is the expected long-range rewards following policy π. The *state-action value function* $Q^\pi : S \times A \to \mathbb{R}$ can be defined similarly. A policy π^* is optimal if $V^{\pi^*}(s) \geq V^{\pi'}(s)$ for all $s \in S$ and all π'. Optimal value functions are denoted V^* and Q^*, and the *optimality equations* are

$$V^*(s) = \max_a \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V^*(s')], \tag{1}$$

$$Q^*(s,a) = \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma \max_{a'} Q^*(s',a')]. \tag{2}$$

A *Relational MDP* is now defined by the tuple $\langle S, A, T, R \rangle$ which is similar to MDP while the states and actions are in relational form [17]. The basic ingredients of it are a set of relational predicates $\mathcal{P}$, a set of action predicates $\mathcal{A}$, and a domain of constants $\mathcal{C}$. The Herbrand base $\mathrm{HB}^{\mathcal{P}\cup\mathcal{C}}$ is the set of all ground atoms which can be constructed form $\mathcal{P}$ and $\mathcal{C}$. The set of all states S' then consists of the power set of $\mathrm{HB}^{\mathcal{P}\cup\mathcal{C}}$. However, the illegal states must be excluded by an implicit logical background theory `BK`. Thus the state space S of the RMDP, denoted as $S^{\mathcal{P}\cup\mathcal{C}}$ in this paper, is $\{s | s \in S', \mathtt{BK} \models s\}$. Similarly, an action space A can be constructed based on $\mathcal{A}$ and $\mathcal{C}$.

3 Logical Markov Decision Processes with Negation

Because the states of the RMDP are in ground level, the size of the state space grows exponentially. As an example of blocks world, based on the relational predicates $\{\mathtt{on/2}, \mathtt{clear/1}\}$, the action predicates $\{\mathtt{move/2}\}$, and the domain $\{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}, \mathtt{e}, \mathtt{floor}\}$, the blocks world containing 5 blocks has 501 legal states. So applying RRL in relational domains must abstract over the ground level.

Definition 1 (Abstract State). *An abstract state s is a fomula $\{s^\top, \mathtt{not}\, s_1^\perp, \cdots, \mathtt{not}\, s_n^\perp\}$, which contains one conjunction $s^\top$ and several ($n \geq 0$) not-conjunctions $\mathtt{not}\, s_i^\perp (i = 1, \cdots, n)$, and holds that $a \notin s_i^\perp$ for any atom $a \in s^\top$. $s^\top$ is called* fit conjunction*, and $s_i^\perp$ are called* unfit conjunctions*.*

An abstract state represents a set of states. From the semantic of `not`, we know that an abstract state $s = \{s^\top, \mathtt{not}\, s_1^\top, \cdots, \mathtt{not}\, s_n^\perp\}$ reperesents a state $\bar{s}$, denoted as $\bar{s} \preceq_\theta s$, iff:

> *There exists θ such that $\bar{s} \preceq_\theta s^\top$ and doesn't exist φ such that $\bar{s} \preceq_\varphi [s^\top \cup s_i^\perp]$ for any $i (i{=}1, \cdots, n)$.*

It means that the state $\bar{s}$ must fit conjunction $s^\top$ and can't fit conjunctions $s_i^\perp$. Use $\bar{S}(s)$ denotes the interpretations of an abstract state s.

The goal abstract state of tasks can be got easily. For one, the goal abstract state of the *stack* task of blocks world (putting all blocks on one stack) is

$\langle s_* \rangle = \{\{\texttt{on(A,B)}, \texttt{clear(A)}\}, \texttt{not}\{\texttt{on(C,D)}, \texttt{clear(C)}, \texttt{C} \neq \texttt{A}\}\}$.

Here, we use "$\neq$" as the only constraint of variables. Note that the number of blocks is not specified. The ground state $\bar{s} = \{\texttt{on(a,b)}, \texttt{on(b,floor)}, \texttt{clear(a)}, \texttt{clear(floor)}\}$ containing two blocks is an interpretation of $\langle s_* \rangle$.

To represent un-goal states, two abstract states can be generated from the goal abstract state $\langle s_* \rangle$:

$\langle s_1 \rangle = \{\varnothing, \texttt{not}\{\texttt{on(A,B)}, \texttt{clear(A)}\}\}$,
$\langle s_2' \rangle = \{\{\texttt{on(A,B)}, \texttt{clear(A)}, \texttt{on(C,D)}, \texttt{clear(C)}, \texttt{C} \neq \texttt{A}\}, \varnothing\}$.

For abstact state $s = \left\{s^\top,\ \texttt{not}\, s_1^\perp, \cdots, \texttt{not}\, s_n^\perp\right\}$, the *generating method* is:

1. if $s^\top \neq \varnothing$, take fit conjunction $\varnothing$ and unfit conjunction $s^\top$ as the first generated abstract state of s;
2. chose m differents unfit conjunctions to the fit part while other unfit conjunctions reserved as the rest of $(2^n - 1)$ generated abstract states of s.

Theorem 1 (Global Complementarity). *Under the closed world assumption, for an abstract state and all generated abstract states followed by above generating method, they are complementary, that is the interpretations of them cover the whole ground state space $S^{\mathcal{P}\cup\mathcal{C}}$, and the intersection of each two interpretations of them is an empty set.*

Proof. The closed world assumption has an effect only on the ground state space $S^{\mathcal{P}\cup\mathcal{C}}$. So each ground state $\bar{s} \in S^{\mathcal{P}\cup\mathcal{C}}$ contains all the true ground atoms of that time. Given a state $\bar{s}$, for an abstract state $s_0 = \{s^\top, \texttt{not}\, s_1^\perp, \cdots, s_n^\perp\}$, each conjunction of $s^\top$ and $s_i^\perp$ must have a true or false value. If $s^\top (\neq \varnothing)$ is false then the first generated abstract state s_1 is true and all the others $(s_0, s_2, \cdots, s_{2^n})$ are false. Then given the true $s^\top$, if all $s_i^\perp$ are false, then s_0 is true, else there must exist only one generated abstract state s_j $(2 \leq j \leq 2^n)$ whose fit conjunction just contains $s^\top$ and all the true $s_i^\perp$, and s_j will be true. We now see $\bigcup_{j=0}^{2^n} \bar{S}(s_j) = S^{\mathcal{P}\cup\mathcal{C}}$ and $\bar{S}(s_j) \cap \bar{S}(s_k) = \varnothing$ $(j \neq k)$, because the interpretations have been restricted to $S^{\mathcal{P}\cup\mathcal{C}}$.

Thus, abstract states $\langle s_* \rangle, \langle s_1 \rangle, \langle s_2' \rangle$ represent all states of stack goal complementally. However, three abstract states is not enough. There must be a method to expand the abstract space and keep the complementarity. The *expanding method* is to add a new conjunction s^+ to certain abstract state $s = \left\{s^\top, \texttt{not}\, s_1^\perp, \cdots, \texttt{not}\, s_n^\perp\right\}$, and replace it with two new abstract states s' and s'':

$$s' = \left\{s^\top \cup s^+, \texttt{not}\, s_1^\perp, \cdots, \texttt{not}\, s_n^\perp\right\}$$
$$s'' = \left\{s^\top, \texttt{not}\, s^+, \texttt{not}\, s_1^\perp, \cdots, \texttt{not}\, s_n^\perp\right\}$$

The former is taking s^+ as a subset of fit conjunction, and the latter is taking s^+ as an unfit conjunction. Obviousely, the new conjunction s^+ must also follow the hold that $a \notin s^\top$ and $a \notin s_i^\perp$ for each atom $a \in s^+$.

Theorem 2 (Local Complementarity). *The two expanded abstract states are complementary for the original abstract state. That is, the interpretaions of the*

two expanded abstract states are the same as that of the original abstract state, and the intersection of them is an empty set.

Proof. If s is true, then there must exist just one abstract state from s' or s'' that is true followed by the value of s^+. So $\bar{S}(s') \cap \bar{S}(s'') = \varnothing$ and $\bar{S}(s') \cup \bar{S}(s'') \supseteq \bar{S}(s)$. No matter s' or s'' is true, s is true. Then $\bar{S}(s') \cup \bar{S}(s'') \subseteq \bar{S}(s)$.

Adding conjunctions $\{\texttt{on(B,E)}\}$, $\{\texttt{on(D,F)}\}$, and $\{\texttt{on(G,H)}, \texttt{clear(G)}, \texttt{G}\neq\texttt{A}, \texttt{G}\neq\texttt{C}\}$ to abstract state $\langle s_2' \rangle$ in turn, the following abstract states are produced:

$\langle s_2 \rangle = \{\{\texttt{on(A,B)}, \texttt{clear(A)}, \texttt{on(C,D)}, \texttt{clear(C)}, \texttt{C} \neq \texttt{A},$
$\texttt{on(B,E)}, \texttt{on(D,F)}, \texttt{on(G,H)}, \texttt{clear(G)}, \texttt{G} \neq \texttt{A}, \texttt{G} \neq \texttt{C}\}, \varnothing\}$
$\langle s_3 \rangle = \{\{\texttt{on(A,B)}, \texttt{clear(A)}, \texttt{on(C,D)}, \texttt{clear(C)}, \texttt{C} \neq \texttt{A}\}, \texttt{not}\{\texttt{on(B,E)}\}\}$
$\langle s_4 \rangle = \{\{\texttt{on(A,B)}, \texttt{clear(A)}, \texttt{on(C,D)}, \texttt{clear(C)}, \texttt{C} \neq \texttt{A}, \texttt{on(B,E)}\}, \texttt{not}\{\texttt{on(D,F)}\}\}$
$\langle s_5 \rangle = \{\{\texttt{on(A,B)}, \texttt{clear(A)}, \texttt{on(C,D)}, \texttt{clear(C)}, \texttt{C} \neq \texttt{A}, \texttt{on(B,E)}, \texttt{on(D,F)}\},$
$\texttt{not}\{\texttt{on(G,H)}, \texttt{clear(G)}, \texttt{G} \neq \texttt{A}, \texttt{G} \neq \texttt{C}\}\}$

Corollary 1 (Complementarity). *Starting from a certain abstract state, and applying the above generating method once and the expanding method several times in turn, the new abstract space is complementary for* $S^{\mathcal{P}\cup\mathcal{C}}$.

Normally, the certain abstract state is the goal abstract state of a task, and the expanding method will not apply to it. These methods give designers a very useful tool to construct the abstract state space in an easy way. The only need for designers is the sequence of the adding conjunctions.

Given an abstract state, the applicable *abstract actions* of it can be gained manually according to the preconditions of actions. For a simple case, consider abstract state $\langle s_2' \rangle$, the applicalbe abstract actions are $\texttt{move(A,C)}$, $\texttt{move(C,A)}$, $\texttt{move(A,floor)}$, and $\texttt{move(C,floor)}$. However, the manual method is very hard for more complex domains. To get abstract actions automatically, the *prototype action* will be introduced next.

Definition 2 (Prototype Action). *An prototype action is an action atom* $\mathring{a}$ *that contains no constants with the precondition and outcome of it, denoted by a tuple* $\langle \mathring{a}, c, P, D, E \rangle$. *Here* $\mathring{a}$ *is the action atom no constances,* c *is the precondition in the form of abstract state,* $P = \{p_1, \cdots, p_m\}$ *is the set of probabilities for* m *different cases by taking the action and holds* $\sum_i p_i = 1$, $D = \{d_1, \cdots, d_m\}$ *is the set of deleting conjunctions for* m *cases, and* $E = \{e_1, \cdots, e_m\}$ *is the set of adding conjunctions for* m *cases.* c, d_i, e_i *all are conjunctions and hold* $vars(D, E) \subseteq vars(c) = vars(\mathring{a})$. *It also can be denoted only by the action atom* $\mathring{a}$, *as* $\mathring{a} : c \xrightarrow{p_i} e_i \setminus d_i$.

This definition implements a kind of probabilistic STRIPS operator. The hold of variables implies that the prototype action can be delegated by the action atom, and the precondition is really a sufficient and necessary condition. Intuitively, prototype actions should be the basic action ways of a domain. Indeed, a prototype action is a super abstract action, and there is only one prototype action $\texttt{move(X,Y,Z)}$ in blocks world. It means moving the block X form block Y

to Z. Consider the prabability 0.1 the agent cannot pickup the block X and 0.1 the block X fall on floor when it is in air, the only prototype action $\langle \mathring{a}_* \rangle$ is

$$\texttt{move(X,Y,Z)}: \left\{ \begin{array}{l} \texttt{on(X,Y)}, \\ \texttt{clear(X)}, \\ \texttt{clear(Z)} \end{array} \right\} \begin{array}{l} \xrightarrow{0.8} \left\{ \begin{array}{l} \texttt{on(X,Z)}, \texttt{clear(Y)}, \\ \texttt{clear(floor)} \end{array} \right\} \setminus \left\{ \begin{array}{l} \texttt{on(X,Y)}, \texttt{clear(Z)}, \\ \texttt{clear(floor)} \end{array} \right\} \\ \xrightarrow{0.1} \{\} \setminus \{\} \\ \xrightarrow{0.1} \left\{ \begin{array}{l} \texttt{on(X,floor)}, \\ \texttt{clear(Y)} \end{array} \right\} \setminus \{\texttt{on(X,Y)}\} \end{array}$$

Definition 3 (Abstract Transition). *An abstract transition is a function* $T : S \times A \times \Theta \times S \to [0, 1]$. $T(s, \mathring{a}, \theta, s')$ *means the probability of changing the abstract state from* s *to* s' *under the effection of a prototype* $\mathring{a}$ *and* θ*, and holds* $\sum_{s'} T(s, \mathring{a}, \theta, s') = 1$. *Here* θ *is a substitution from vars*$(\mathring{a})$ *ot terms*(s) *and holds* $c(\mathring{a})\theta \subseteq s$.

Given an abstract state and a prototype action, all valid substitutions are known. Indeed, a substitution θ effecting on a prototype action $\mathring{a}$ can produce an applicable abstract action $\mathring{a}\theta$ of an abstract state s.

The probabilies of the prototype action are hidden on the abstract transitions. For model-free algorithms, it is enough that only learn the transition probabilites while the probabilites of prototype actions are unknown.

The *abstract reward* and the *abstract policy* can be similarly defined.

Definition 4 (nLMDP). *A Logical Markov Decision Process with Negation (nLMDP) is a tuple* $M = \langle S, A, T, R \rangle$ *where* S *is a given complementary abstact space,* A *is a set of prototype actions,* T *is a finite set of abstract transitions based on* S *and* A*, and* R *is a reward model.*

A nLMDP is based on a given abstract state space and a given set of prototype actions while the valid substitutions can be deterministic. And the state space is complementary, so many RL algorithms based on an underling MDP can be used in nLMDP directly.

4 $\Theta(\lambda)$ Learning

To learn an optimal policy based on the prototype action, there should have a value function for evaluating the triple of an abstract state, a prototype action, and a valid substitution. To distinguish the *state-action-substitution value function* $S \times A \times \Theta \to \mathbb{R}$ from state value function and state-action value function, it is also denoted as Θ. Indeed, Θ just divides the action of Q into a prototype action and a substitution. So the optimality equation for Θ is

$$\Theta^*(s, \mathring{a}, \theta) = \sum_{s'} T(s, \mathring{a}, \theta, s')[R(s, \mathring{a}, \theta, s') + \gamma \max_{\mathring{a}'} \max_{\theta'} \Theta^*(s', \mathring{a}', \theta')], \tag{3}$$

and the relation with V^* is

$$V^*(s) = \max_{\mathring{a}} \max_{\theta} \Theta^*(s, \mathring{a}, \theta). \tag{4}$$

Input: ground state $\bar{s}$, value funtion Θ
Output: abstract state s, substitution φ, prototype action $\mathring{a}$, substitution θ
1: Find the unique abstract state s and a substitution φ such that $\bar{s} \preceq_\varphi s$
2: Find a set of ground actions based on $\bar{s}$ such that $\bar{A} = \{\mathring{a}\psi | \bar{s} \preceq_\psi c(\mathring{a})\}$
3: Select a set of ground actions from $\bar{A}$ based on s and φ such that $\bar{A}' = \{\mathring{a}\psi | (\mathring{a}, \psi) \in A, terms(\mathring{a}\psi) \subseteq terms(s\varphi)\}$
4: Transform $\bar{A}'$ to applicable abstract actions A based on s and φ such that $A = \{(\mathring{a}, \theta) | \mathring{a}\psi \in \bar{A}', \mathring{a}\psi = \mathring{a}\theta\varphi\}$
5: Chose a pair $(\mathring{a}, \theta)$ from A based on Θ and a policy (e.g. softmax)

Fig. 1. Algorithm of automatical action abstracting

The first task of Θ-learning is to get the applicable abstact actions automatically. Fig. 1 shows the algorithm. The agent finds the ground actions $\bar{A}$ based on the precondition of prototype actions firstly, selects valid actions $\bar{A}'$ based on term constraints, transforms to abstract actions A, and chose a pair $(\mathring{a}, \theta)$ based on Θ value function and softmax policy [2]:

$$p((\mathring{a}, \theta)|s) = \frac{e^{\Theta(s,\mathring{a},\theta)/T}}{\sum_{\mathring{a}'} \sum_{\theta'} e^{\Theta(s,\mathring{a}',\theta')/T}}. \tag{5}$$

1: $\Theta \leftarrow \varnothing$
2: **for** each episode **do**
3: Initialize ground state $\bar{s}$, $\mathcal{E} \leftarrow \varnothing$
4: Get (s, φ) and $(\mathring{a}, \theta)$ by calling algorithm 1 based on $\bar{s}$
5: **while** s is not the goal abstract state **do**
6: Take action $\mathring{a}\theta\varphi$, observe reward r and successor state $\bar{s}'$ ($\bar{s}' = [\bar{s} \setminus d_i(\mathring{a})\theta\varphi] \cup e_i(\mathring{a})\theta\varphi$ with probability $p_i(\mathring{a})$)
7: Get (s', φ') and $(\mathring{a}', \theta')$ by calling algorithem 1 based on $\bar{s}'$
8: **if** $\Theta(s, \mathring{a}, \theta)$ is in Θ **then**
9: $\Theta^0 \leftarrow \Theta(s, \mathring{a}, \theta)$
10: **else**
11: $\Theta^0 \leftarrow 0$, add $\Theta(s, \mathring{a}, \theta) = 0$ to Θ
12: add $e(s, \mathring{a}, \theta) = 0$ to $\mathcal{E}$
13: **if** $\Theta(s', \mathring{a}', \theta')$ is in Θ **then** $\Theta' \leftarrow \Theta(s', \mathring{a}', \theta')$ **else**
14: $\Theta' \leftarrow 0$
15: $\delta \leftarrow r + \gamma\Theta' - \Theta^0$
16: $e(s, \mathring{a}, \theta) \leftarrow e(s, \mathring{a}, \theta) + 1$
17: **for** each triple $(s, \mathring{a}, \theta) \in \mathcal{E}$ **do**
18: $\Theta(s, \mathring{a}, \theta) \leftarrow \Theta(s, \mathring{a}, \theta) + \alpha\delta e(s, \mathring{a}, \theta)$
19: $e(s, \mathring{a}, \theta) \leftarrow \gamma\lambda e(s, \mathring{a}, \theta)$
20: $s \leftarrow s', \varphi \leftarrow \varphi', \mathring{a} \leftarrow \mathring{a}', \theta \leftarrow \theta'$

Fig. 2. $\Theta(\lambda)$-learning

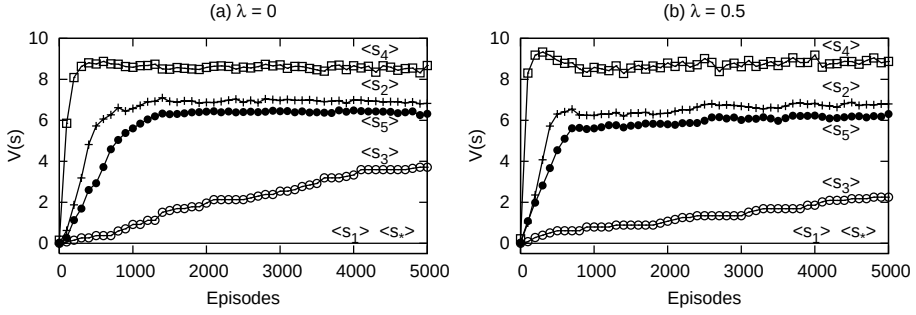


Fig. 3. Experiment results of $\Theta(\lambda)$-learning with (a) $\lambda = 0$ and (b) $\lambda = 0.5$

The second task of Θ-learning is to evaluate the value fungion Θ through iterations. Fig. 2 shows $\Theta(\lambda)$-learning method, which uses an eligibility trace decay parameter λ similar with Sarsa(λ) [2]. To prevent invalid substitutions from occuring in Θ which is initialized with an empty set in line 1, valid substitutions are added in real-time in line 11. In line 17-19, triples of the episode are updated, and $\mathcal{E}$ is used to record the triples and initialized with an empty set in the beginning of each episode in line 3.

5 Experiments

We implemented Θ-learning using the system of GNU Prolog 1.2.18 for the stack task of blocks world. The blocks world is the prototypical toy domain requiring relational representations and has been experimented usually in RRL. Only two predicates `on/2` and `clear/1` are used, and other predicates such as `block/1`, `floor/1`, `above/2`, `height/2` reported by [14,3] are unnecessary in our model and methods. Six abstract states $\langle s_* \rangle$, $\langle s_1 \rangle - \langle s_5 \rangle$ and the only prototype action $\langle \mathring{a}_* \rangle$ of the task have been listed in section 3. Because `not` is used, the abstract state space is complementary and constructed in an easy way by using the generating and expanding methods. The agent gets reward 10 when it reaches the goal abstract state $\langle s_* \rangle$, while 0 when it reaches other states.

The discount factor γ was 0.9, the learning rate α was 0.015, and the temperature T of the policy softmax was decreased by 0.996 each episode starting with 1.0. 500 initial states of blocks world were generated randomly for 6 to 10 blocks (more than 58 million ground states) using the procedure described by [18], which guarantees that the initial states are followed by the background theory of blocks world. The initial ground state of each episode was selected randomly from them. Fig. 3 shows the evaluated optimal state value funtion of each abstract state for $\lambda = 0$, and $\lambda = 0.5$. The total running time for all 5000 episodes was less than 2 minutes on an 1 GHz Linux machine estimated using GNU Prolog's build-in predicate `user_time(Time)`.

It shows clearly that $\lambda = 0.5$ learns faster than $\lambda = 0$ in the begining, and both converge. The reason is that the non-zero reward is sparse, so bigger λ can

effect more triples $(s, \mathring{a}, \theta)$ when a non-zero reward is received. It shows that using an eligibility trace parameter is a feasible method for solving such kinds of tasks. Episode ends when the goal abstract state $\langle s_* \rangle$ reached, so $V(s_*)$ is always 0. Abstract state $\langle s_1 \rangle$ represents a blocks world containing 0 blocks, so $V(s_1)$ is always 0, too. Abstract state $\langle s_3 \rangle$ represents all blocks on the floor, for the stack task, it will be visited seldom and all applicable abstract actions are equivalence, so the speed of convergence of $\langle s_3 \rangle$ is very slow. The learned abstract actions for $\langle s_2 \rangle, \langle s_4 \rangle, \langle s_5 \rangle$ are all `move(C,D,A)`.

6 Related Work

Recently, there has been an increased and significant interests in using rich relational representations for modeling and learning MDPs. For the modeling MDP, Van Otterlo [11,17] gave the definition of the RMDP in the ground level and defined the *CARCASS* which is an RL-based abstraction for RMDPs, and several abstract models have been proposed. Kersting and De Raedt [10,14] defined the *Logical MDP* (LOMDP) which is an abstraction over RMDPs by using variables in states and actions, and use Q-learning and TD(0) on the fixed abstract model. Kersting and Van Otterlo [19] introduced ReBel which uses a constraint logic language to regress preceding abstract states automated based on LOMDP for model-based learning. The main differences between LOMDPs and our nLMDPs are the negation introduced in nLMDPs, the complementarity of abstract states of nLMDPs, and the definition of the action space. Butilier et al. [9] gave *symbolic dynamic programming* (SDP) which is a situation calculus abstraction over an RMDP, but it is expressive logic that did not permit an efficient computational algorithm. Morales [12] introduced an *a priori* abstraction of RMDPs based on separate state and action spaces and uses Q-learning over the abstraction. Guestrin et al. [13] used probabilistic relational models and modelled class-based value functions assuming fixed relations.

For other model-free approaches, Džeroski et al. [3], Lecoeuche [20], Driessens & Ramon [4], Gärtner et al. [5], Cole et al. [6], Yoon et al. [7], and Fern et al. [8] have studied different relational learners for function approximation.

7 Conclusions

In nLMDPs, the logical negation is first introduced under the closed world assumption. So a complementary abstract state space can be constructed in an easy way by applying the generating method once and the expanding method several times in turn until the size of it is suitable for learning. Based on the complementary abstract state space and the definition of prototype action, the nLMDP is proposed, and a model-free Θ-learing method is implemented for evaluating the state-action-substitution value funcion. Experiments perform well and show convergence results.

Future works could include: constructing abstract state space automatically, extending the language to gain stronger expressive power, and other algorithms based on the nLMDP or its extensions.

References

1. Kaelbling, L., Littman, M., Moore, A.: Reinforcement learning: A survey. Journal of Artificial Intelligence Research **4** (1996) 237–285
2. Sutton, R., Barto, A.: Reinforcement Learning: An Introduction. The MIT Press (1998)
3. Džeroski, S., De Raedt, L., Driessens, K.: Relational reinforcement learning. Machine Learning **43** (2001) 7–52
4. Driessens, K., Ramon, J.: Relational instrance based regression for relational reinforcement learning. In: ICML-2003. (2003)
5. Gärtner, T., Driessens, K., Ramon, J.: Graph kernels and gaussian processes for relational reinforcement learning. In: ILP'03. (2003)
6. Cole, J., Lloyd, K., Ng, K.: Symbolic learning for adaptive agents. In: The Annual Partner Conference, Smart Internet Technology Cooperative Research Centre. (2003)
7. Yoon, S., Fern, A., Givan, R.: Inductive policy selection for first-order MDPs. In: UAI'02. (2002)
8. Fern, A., Yoon, S., Givan, R.: Approximate policy iteration with a policy language bias. In: NIPS'03. (2003)
9. Boutilier, C., Reiter, R., Price, B.: Symbolic dynamic programming for first-order MDPs. In: Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-01). (2001) 690–700
10. Kersting, K., De Raedt, L.: Logical markov decision programs. In: IJCAI'03 Workshop on Learning Statistical Models of Relational Data. (2003)
11. Van Otterlo, M.: Reinforcement learning for relational MDPs. In: Machine Learning Conference of Belgium and the Netherlands (BeNeLearn'04). (2004)
12. Morales, E.: Scaling up reinforcement learning with a relational representation. In: Proceedings of the Workshop on Adaptability in Multi-agent Systems at AORC'03, Sydney. (2003)
13. Guestrin, C., Koller, D., Gearhart, C., Kanodia, N.: Generalizing plans to new environments in relational MDPs. In: IJCAI'03. (2003)
14. Kersting, K., De Raedt, L.: Logical markov decision programs and the convergence of logical TD(λ). In: Fourteenth International Conference on Inductive Logic Programming (ILP'04). (2004) 180–197
15. Neinhuys-Cheng, S.H., de Wolf, R.: Foundations of Inductive Logic Programming, vol. 1228 of Lecture Notes in Artifical Intelligence. Springer-Verlag (1997)
16. Clark, K.: Negation as failure. In: Logic and Data Bases. (1977) 293–322
17. Van Otterlo, M., Kersting, K.: Challenges for relational reinforcement learning. In: ICML'04 Workshop on Relational Reinforcement Learning. (2004)
18. Slaney, J., Thiébaux, S.: Blocks world revisited. Artificial Intelligence **125** (2001) 119–153
19. Kersting, K., Van Otterlo, M., De Raedt, L.: Bellman goes relational. In: ICML'04. (2004)
20. Lecoeuche, R.: Learning optimal dialogue management rules by using reinforcement learning and inductive logic programming. In: Proc. of the North American Chapter of the Association for Computational Linguistics (NAACL). (2001)

Mobile Agent Routing Based on a Two-Stage Optimization Model and a Hybrid Evolutionary Algorithm in Wireless Sensor Networks

Shaojun Yang, Rui Huang, and Haoshan Shi

School of Electronics and Information, Northwestern Polytechnical University
Xi'an, Shaanxi, 710072, P.R. China

Abstract. A new two-stage optimization model for mobile agent routing in the wireless sensor network is presented as a integer linear programming problem based on the virtual connection topology sub graph. In the model, the solution is presented by two subpaths instead of the closed path which is usually applied in many path search methods to keep balance between computation cost and accuracy. A hybrid technique of genetic algorithm (GA) integrated with discrete particle swarm optimization (PSO), GAPSO is designed to solve the problem. Simulation experiments with different sizes and distributions of nodes show the effectiveness of the new model and GAPSO algorithm.

1 Introduction

The emerging Wireless Sensor Networks (WSNs) composed of tiny, densely deployed sensors hold the promise of unprecedented ability to monitor and instrument the physical phenomenon [1]. In a typical wireless sensor network, a set of nodes with constrained resource such as limited battery power, limited communication and computation abilities have presented unique challenges including fault tolerance, scalability, reliability and power consumption in network design and Collaborative Signal and Information Processing (CSIP) [1, 2, 3]. The traditional data processing and fusion architecture is based on the client/server computing model, which may have negative effects on the security of data due to the multi-source and multi-path model. To overcome these limitations, mobile agent paradigm with wide applications in various areas [4,5] has been introduced as a promising solution. Qi *et al* [6, 7] proposed the mobile agent based distributed sensor network for scalable and energy-efficient data aggregation. In the model, mobile agent with partially integrated results generated from the sensor nodes visited migrates among the nodes of interest according to some path and thus the progressive accuracy can be obtained. The mobile agent based model is suitable for collaborative processing and recently, the interests in it has been on the increase.

One of the crucial problems in the mobile agent based computing paradigm is how to plan the itinerary (or route) for a mobile agent in order to achieve progressive fusion accuracy at minimal expense of resources. The problem of mobile agent routing has been investigated in [8, 9, 10]. Two simple algorithms,

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 938–947, 2006.

Global Closest First (GCF) and Local Closest First (LCF), have been proposed [11]. LCF searches for the next node with the shortest distance to the current node, while GCF searches for the next closest node to some given node. With the increase in the size of network and the number of nodes, the performance of these two heuristics would deteriorates. Evolutionary computing which has been widely used in communication applications [12] may be a more appealing option. Refs [13] and [14] applied genetic algorithm (GA) to design the route for mobile agent and obtained satisfactory results. For easy implementation, most methods assume that each sensor node should be visited no more than once in the path. In fact, the assumption is unnecessary since a feasible solution is a closed path returning to its starting point with required accuracy. However, search in this solution space is more difficult.

In this paper, we focus our discussion on the routing problem for the mobile agent in the WSNs. The virtual connection topology sub graph (VCTSG) including sink nodes, key relay nodes and active sensor nodes is first constructed. Then, we formulate the mobile agent routing problem as an integer linear programming and propose a two-stage optimization model, where the visiting path is divided into two subpaths, one for gathering and processing data and the other for returning the results, and both of them should not visit duplicated nodes. GAPSO, a hybrid algorithm based on GA and Particle Swarm Optimization (PSO) is presented to solve the mathematic model and program the itinerary.

2 Mobile Agent Routing in WSNs

2.1 Mobile Agent Based WSN Model

The mobile agent based WSN typically consists of the sensor nodes, observer nodes and mobile agents. The sensor nodes are usually scattered in a sensing field. Each of sensor nodes locally carries out the physical sensing of environmental phenomena and detects events of interest. Compared with sensor nodes, observer node is usually assigned more resources for computation, communication and store. It not only acts as a commander or sink to indicate interests (or queries) to the network and receive event reports about these queries, it also communicates with the external networks as a relay or gateway. Mobile agent is a special software module dispatched by the observer node. It can visit and process data in sensor nodes along a predetermined itinerary.

2.2 Problem Formulation

Let V denote the set of all the nodes with maximum communication range $R_{\max}$. s is the observe node and T is the set of sensor nodes that have detected the potential targets. Usually, the transmission cost c_{ij} from node i to node j $(i, j \in V)$ involves the path loss and time consumption. In this paper, we mainly consider the effects of path loss which is defined as the difference (in dB) between the effective transmitted power and received power according to the Friis free space propagation model [15]. Then c_{ij} can be calculated as

$$c_{ij} = \begin{cases} 10\log\left((4\pi)^2 R_{ij}^2/\lambda^2\right), & R_{ij} \le R_{\max} \\ \infty, & R_{ij} > R_{\max} \end{cases} \tag{1}$$

where R_{ij} is the distance between nodes i and j, λ is the ratio of light speed and carrier frequency.

When a mobile agent is dispatched from the observe node to collect and integrate data, low transmission consumption and high detection accuracy are always the main goals for itinerary programming. Let x_{ij} indicate the link state between nodes i and j. $x_{ij} = 1$ means the link is active and $x_{ij} = 0$ otherwise. We formulate the problem as an integer linear programming to search a path with minimal cost and required accuracy:

$$\min \sum_{i,j \in V, i \neq j} c_{ij} x_{ij} \tag{2}$$

subject to:

$$\begin{aligned} \sum_{i \in W, j \in \bar{W}} x_{ij} \ge 1, \quad \{s\} \subset W \subset V, \bar{W} \cap T \neq \emptyset \\ \sum_{i \in U, j \in \bar{U}} x_{ij} \ge 1, \quad T \subset U \subset V, \bar{U} \cap \{s\} \neq \emptyset \end{aligned} \tag{3}$$

$$\sum_{j \in T'} \mathrm{SE}(j) \ge \mathrm{SE}_T, \quad T' = \{j x_{ij} | \forall i \in V, \forall j \in T\} \tag{4}$$

where $x_{ij} \in \{0, 1\}$, $W \cup \bar{W} = V$, $U \cup \bar{U} = V$, $\mathrm{SE}(j) \ge 0$ is the signal energy detected by the node j, and SE_T is the signal threshold.

As the connectivity constraint, (3) means the mobile agent should move from the observe node to the sensor nodes which have detected the potential targets and then return back to the observer. Since the targets' signal energy received by the sensor nodes can be used by the mobile agent for data fusion, the mobile agent always tries to accumulate as much signal energy as possible for accurate decision in target detection. Considering the fact of limited resource, energy constraint (4) states the mobile agent should collect enough amount of signal energy larger than a preset signal threshold.

2.3 Suboptimal Model

From the above mathematic model, the constraint of visiting each node besides the observer no more than once is unnecessary, but it has been adopted in many heuristic search methods as a suboptimal model to reduce the solution space for the sake of easy implementation. Considering the balance between computational cost and accuracy, we present a two-stage suboptimal model, where the path is divided into two, one called the "integrating subpath" for migrating from the observer to sensor nodes to collect and integrate data and the other called the "returning subpath" for turning back to the beginning of the path. Each subpath should be composed of different nodes, while the different subpath could have the same nodes. For the two subproblems, we define the same target function as (2) with different constraints as follows

1. in integrating subpath

$$\begin{aligned} \sum_{i\in W, j\in \bar{W}} x_{ij} \geq 1, \quad \{s\} \subset W \subset V, \bar{W} \cap T \neq \emptyset \\ \sum_{j\in V} x_{ij} \leq 1, \quad \sum_{i\in V} x_{ij} \leq 1 \end{aligned} \tag{5}$$

$$\sum_{j\in T'} \mathrm{SE}(j) \geq \mathrm{SE}_T, \quad T' = \{j x_{ij} | \forall i \in V, \forall j \in T\} \tag{6}$$

2. in returning subpath

$$\begin{aligned} \sum_{i\in U, j\in \bar{U}} x_{ij} \geq 1, \quad T \subset U \subset V, \bar{U} \cap \{s\} \neq \emptyset \\ \sum_{j\in V} x_{ij} \leq 1, \quad \sum_{i\in V} x_{ij} \leq 1. \end{aligned} \tag{7}$$

3 Mobile Agent Routing Based on the Hybrid Algorithm of GA and PSO (GAPSO)

3.1 PSO and Its Discrete Versions

PSO is proposed according to the observations of the social behavior of animals [16, 17]. In PSO, a swarm move around in a D-dimensional search space. During the search process the particle successively adjusts its position $\mathbf{x}_i = (x_{i1}, \ldots, x_{id}, \ldots, x_{iD})$ toward the global optimum according to the best position encountered by itself denoted as $\mathbf{p}_i = (p_{i1}, \ldots, p_{id}, \ldots, p_{iD})$ and the best position encountered by the whole swarm denoted as $\mathbf{p}_g = (p_{g1}, \ldots, p_{gd}, \ldots, p_{gD})$ after each iteration. Let its velocity $\mathbf{v}_i = (v_{i1}, \ldots, v_{id}, \ldots, v_{iD})$, the PSO algorithm is described as follows [18]:

$$v_{id} = w v_{id} + c_1 r_1 (p_{id} - x_{id}) + c_2 r_2 (p_{gd} - x_{id}) \tag{8}$$

$$x_{id} = x_{id} + v_{id} \tag{9}$$

where c_1 and c_2 are positive constants called acceleration coefficients, r_1 and r_2 are two random numbers in the range [0, 1], and w is an inertia factor. In addition, the velocities of the particles are clamped to $[V_{\min}, V_{\max}]^D$.

Since the PSO technique described above is the real valued PSO, it should be extended to deal with many discrete optimization problems which require the ordering or arranging of discrete elements, eg. the mobile agent routing problem. A popular solution is to keep the velocity update formula (8) unchanged but compute the actual new position component to be 1 or 0 with a probability obtained by applying a sigmoid transformation $S(v_{id}) = 1/(1 + \exp(-v_{id}))$ to the velocity component [19]. Another method views the position vectors of the particles as probabilities and employing roulette wheel selection for discretization [20]. The two techniques both extend the real valued PSO to its binary

version, namely Binary PSO (BPSO), by a simple discretization of the values of velocities or positions of particles. A more general discrete PSO is proposed in [21] to solve the Traveling Salesman Problem. The method redefines the six basic mathematical objects and operations in the discrete space.

3.2 Virtual Connection Topology Sub Graph (VCTSG)

As a NP complete problem, the mobile agent routing can be solved by the global optimization methods which need the information of network topology and visiting cost of nodes in hand. However, it is relatively difficult to satisfy the requirement for WSNs composed of a large amount of nodes with limited resources.

In [22], we have realized the mobile agent entity and environment based on Directed Diffusion (DD) which is a popular mechanism for data dissemination in WSNs [23, 24]. In the paper, the VCTSG including the observer node, key relay nodes and active sensor nodes is further constructed via data flooding and path reinforcement of DD to handle the problems about how to collect the information of topology and nodes, and to cut down the complexity of solution space.

3.3 GAPSO for Mobile Agent Routing

GAPSO is the genetic operations combined with the particle's position update of discrete PSO. The detailed algorithm design is presented as following.

Fitness Functions. As it is relatively complex to search the solution in the form of link state matrix $\mathbf{X}$, the solution is directly presented by the node sequence P being visited by mobile agent. To facilitate the design of GAPSO algorithm, we respectively define the fitness functions for the integrating subpath and returning subpath as

$$f_{\text{int}}(P) = A - (J(P) + g(P)) \tag{10}$$

$$f_{\text{ret}}(P) = A - J(P) \tag{11}$$

where A is pre-specified to keep the fitness values nonnegative, $J(P)$ is the path loss when mobile agent visiting nodes in P, $g(P)$ is the penalty function for the constraint (6) in the integrating subpath and is defined by

$$g(P) = \begin{cases} 0, & \sum_{j\in P} \text{SE}(j) \geq \text{SE}_T \\ \beta \cdot (\sum_{j\in P} \text{SE}(j) - \text{SE}_T), & \sum_{j\in P} \text{SE}(j) < \text{SE}_T \end{cases} \tag{12}$$

where β is a properly selected penalty coefficient. In (10), $J(P)$ and $g(P)$ are assumed to be normalized to appropriately reflect the contribution by the path loss and signal gain. Intuitively, the fitness function prefers paths with required amount of signal energies at less transmission cost. The other constraints would be satisfied via encoding.

Encoding. For GA, we adopt the two-level genetic encoding scheme where the first level is a numerical encoding of the nodes sequence in the order of nodes being visited by mobile agent and the second level is a binary encoding of the visit status sequence in the same visiting order (1 and 0 in the k-th position indicate the k-th node is visited and unvisited, respectively). For PSO, only the first level encoding scheme of GA is adopted.

Initialization. The initial population is randomly generated with some constraints. For the integrating subpath, the feasible individuals are those whose numerical strings all star from node s and the first bits in the binary strings are always set 1 with the amount of signal energy collected by mobile agent satisfying the (6). As for the returning subpath, individuals are feasible if their numerical string are from the last visited node in the integrating subpath to node s and the corresponding bits in the binary string are 1. PSO takes the feasible numerical strings as its population.

Crossover. Crossover operation is to combine two individuals to create offsprings which may be better than the parents if they inherit the best traits of both parents. The Roulette selection is used to select the parents for crossover. For a numerical string, crossover operation usually leads to some duplicated nodes in the resulting string and fails to guarantee that each node appears exactly once. Therefore, we use the discrete PSO method [21] to generate the new numerical string. The two-point crossover is applied to the binary string. A high crossover probability is beneficial to search new solution space.

Mutation. Mutation is an important part of the genetic search intended to prevent the population from stagnating at any local optima and thus to keep the variety of the solution. Two mutation techniques, two-point mutation and inversion, are adopted to both levels of strings. Because a high probability for mutation operation may tend to make the search into a primitive random search, a relatively small mutation rate is favorable.

Solution optimization in returning subpath. In each iteration, the population updated by GA is again optimized using PSO, where the particles are presented as the connectivity matrixes and the method of [19] is applied to discretization. Ref [25] reported that PSO can achieve faster convergence speed when using the encoding and discretization formats in the case of searching a path with the least cost and the known target node.

4 Experiments

To assess the performance of GAPSO based mobile agent routing algorithm, we compare the search results of it with those obtained from the LCF, GA and BPSO methods. At the same time, the comparisons between the two optimization models, namely model-I searching one closed path without any duplicated

nodes except for the observer on it and model-II searching two subpaths with constraint that no same nodes should be visited more than once in each subpath, are carried out. Therefore, there are totally seven methods to be compared: GA for model-I (GAM-I) and model-II (GAM-II), GAPSO for model-I (GAPSOM-I) and model-II (GAPSOM-II), BPSO for model-I (BPSOM-I) and model-II (BPSOM-II), LCF for model-II (LCFM-II). Considering the fact that as a greedy algorithm, LCF may fail to find a closed path in model-I due to the limitation of transmission range of each node, we only apply it to model-II.

The nodes are randomly deployed in a 1000×1000 square. A series of experimental networks with different numbers of sensor nodes varying from 10 to 100 are created and each network size includes 10 random scenes. $R_{\max}$ is determined by the area of sensing field and the number of nodes [26].

In BPSO and GAPSO, the related parameters can be chosen as following: the population size is set to 50; the maximum generation is 500; the crossover rate and mutation rate are set 0.95 and 0.25, respectively; the inertia factor w is gradually decreased from 1.2 towards 0.4; the acceleration coefficients $c1$ and $c2$ are 2.0. The same parameters are also adopted in GA. All methods use the same fitness functions, and the penalty coefficient $\beta = 0.6$. For the evolutionary algorithms, the iterative procedure will stop if the relative difference between maximum and minimum of fitness values is less than 10^{-6} after the population update. Due to the inherent randomness, each of the algorithms except for LCFM-II run 5 times in every scene and select the best solutions for comparison.

Fig. 1 depicts the comparison of the ratio of path loss and information gain of the seven methods for the different node sizes. From the figure, we can see that the ratios generally decrease with the increase in the network density. LCFM-II produces the worst results in these methods, as it is based on local optimization strategy. Compared with it, the corresponding ratios obtained from the rest six methods in various network sizes show less differences in the current scales. Fig. 2 gives a clearer picture for performance of the six methods. It can be seen that the optimization methods based on model II achieve better results than those based on model I. In both models, BPSO performs worst and GAPSO generally competes well with GA after the number of nodes grows up to 40.

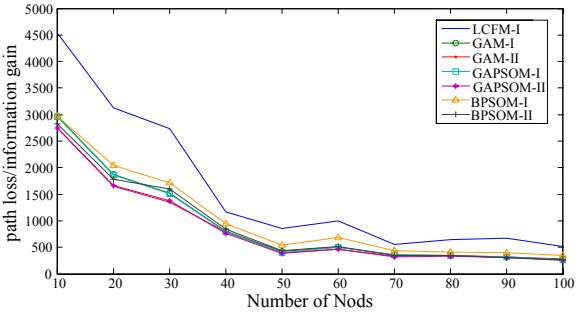


Fig. 1. Ratios of path loss and information gain of seven methods

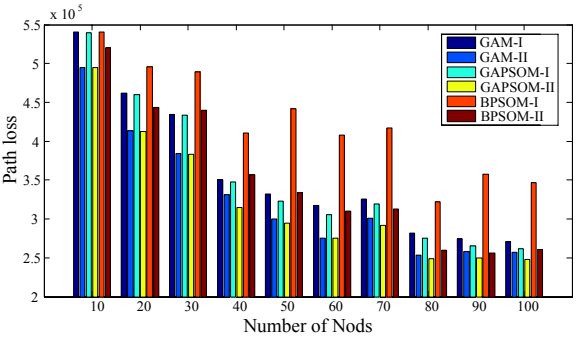


Fig. 2. Performance comparison of three evolutionary algorithms in two models

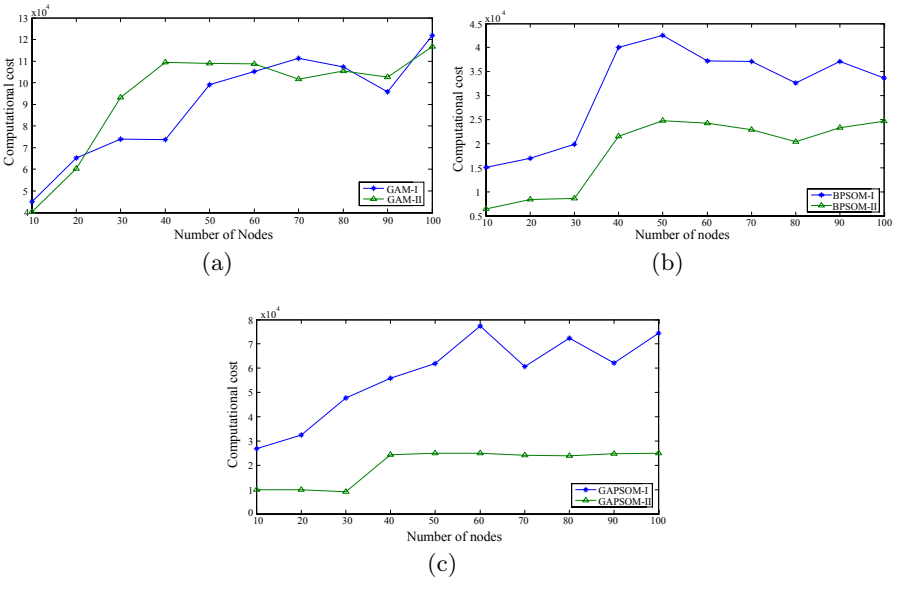


Fig. 3. Comparison of computational cost between two models

In addition, we compare the two models' amount of computation which can be measured by the number of evaluation times of the fitness function. Fig. 3 gives the comparison of average computation cost between the two models using the three evolutionary algorithms for the networks with ten different numbers of nodes. It can be clearly seen that compared with GAPSO and BPSO, the selection of models has less effect on the computation amount of GA. For the two former methods, model II can significantly alleviate the computational burden which shows smoother changes when the network size bigger than 40.

5 Conclusion

The paper proposes a two-stage optimization model for mobile agent routing in WSNs where the procedure of itinerary programming is respectively implemented by planning two subpaths instead of one closed path. At the same time, a hybrid algorithm of GA and discrete PSO is presented to solve the integer linear programming problem. Experiment results demonstrate better performance of model II, especially when the number of nodes becomes larger. Compared with GA and BPSO, GAPSO which integrates discrete PSO into the genetic crossover operation achieves better results.

Acknowledgement

The series of research work dealt in this paper are supported in partially by National Natural Science Fund and PhD research Fund of the Ministry of Education in China.

References

1. Akyildiz, I., W. Su, Y.S., Cayirci, E.: Wireless sensor networks: A survey. Computer Networks **38** (2002) 393–422
2. Estrin, D., Govindan, R., Heidemann, J., et al: Next century challenges: Scalable coordination in sensor networks. In: Proc. ACM/IEEE International Conference on Mobile Computing and Networking, Seattle, Washington, USA (1999) 263–270
3. Chong, C.Y., Kumar, S.: Sensor networks evolution, opportunities, and challenge. In: Proc. IEEE. Volume 91. (2003) 1247–1256
4. Tripathi, A., Ahmed, T., Pathak, S., et al: Paradigms for mobile agent based active monitoring of network systems. In: IEEE/IFIP Network Operations and Management Symposium. (2002) 65–78
5. Wang, T., Guan, S., Chan, T.: Integrity protection for code-on-demand mobile agents in e-commerce. Journal of System and Software **60** (2002) 211–221
6. Qi, H., Xu, Y.: Mobile-agent-based collaborative signal and information processing in sensor networks. Proc. IEEE **91** (2003) 1172–1183
7. Qi, H., Iyengar, S., Chakrabarty, K.: Multiresolution data integration using mobile agents in distributed sensor networks. IEEE Trans. Syst., Man, Cybern. **31** (2001) 383–391
8. Migas, N., Buchanan, W.J., McAartney, K.A.: Mobile agents for routing, topology discovery, and automatic network reconfiguration in ad-hoc networks. In: Proc. 10th IEEE International Conference on Engineering of Computer-Based Systems, Huntsville, AL, USA (2003) 200–206
9. Lu, S., Xu, C.: A formal framework for agent itinerary specification, security reasoning and logic analysis. In: Proc. 3rd International Workshop on Mobile Distributed Computing, Columbus, Ohio, USA (2005) 580–586
10. Avramopoulos, I.C., Anagnostou, M.E.: Optimal component configuration and component routing. IEEE Trans. Mobile Comput. **1** (2002) 303–312
11. Qi, H., Wang, F.: Optimal itinerary analysis for mobile agents in ad hoc wireless sensor networks. In: Proc. 13th International Conference on Wireless Communications. Volume 1., Calgary, Canada (2001) 147–153

12. Wang, L.P., ed.: Soft Computing in Communications, Springer, Berlin Heidelberg New York (2003)
13. Selamat, A., Omatu, S.: Analysis on route selection by mobile agents using genetic algorithm. SICE 2003 Annual Conference **2** (2003) 2088–2093
14. Wu, Q., Rao, N.S., Barhen, J., et al: On computing mobile agent routes for data fusion in distributed sensor networks. IEEE Trans. Knowledge Data Eng. **16** (2004) 740–753
15. Rappaport, T.: Wireless Communications Principles and Practice, Second Edition. Publishing House of Electronics Industry, Beijing, China (2004)
16. Eberhan, R.C., Kennedy, J.: A new optimizer using particle swarm theory. In: Proc. 6th International Symposium on Micromachine and Human Science, Nagoya, Japan (1995) 39–43
17. Kennedy, J., Eberhart, R.C.: Particle swarm optimization. In: Proc. IEEE International Conference on Neural Networks, Perth, Australia (1995) 1942–1948
18. Shi, Y., Eberhart, R.C.: Empirical study of particle swarm optimization. In: Proc. Congress on Evolutionary Computation, Piscataway, USA (1999) 1945–1950
19. Kennedy, J., Eberhart, R.C.: A discrete binary version of the particle swarm algorithm. In: Proc. IEEE International Conference on Systems, Man, and Cybernetics, Piscataway, New Jersey, USA (1997) 4104–4109
20. Agrafiotis, D.K., Cedeño, W.: Feature selection for qsar and qspr using binary particle swarms. J. Med. Chem. **45** (2002) 1098–1107
21. Clerc, M.: Discrete particle swarm optimization illustrated by the traveling salesman problem. http://clerc.maurice.free.fr/pso (2000)
22. Yang, S., Shi, H., Huang, R.: Study of spatio-temporal information integration framework based on directed diffusion and mobile-agen for wireless sensor networks. Journal of Electronics and Information Technology **27** (2005) 1994–1999
23. Intanagonwiwat, C., Govinda, R., Estrin, D., et al: Directed diffusion for wireless sensor networking. IEEE/ACM Transactions on Networking **11** (2002) 2–16
24. Silva, F., Heidemann, J., Govindan, R., et al: Directed diffusion. Technical Report ISI-TR-2004-586, USC/Information Sciences Institute (2004)
25. Yuan, P., Ji, C., Zhang, Y., et al: Optimal multicast routing in wireless ad hoc sensor networks. In: Proc. 2004 IEEE International Conference on Networking, Sensing and Control, Taipei, Taiwan (2004) 367–371
26. Koskinen, H.: Connectivity and reliability in ad hoc networks. Master's thesis, Department of Electrical and Communications Engineering, Helsinki University of Technology, Helsinki, Finland (2003)

Solving Uncertain Markov Decision Problems: An Interval-Based Method

Shulin Cui [1,3], Jigui Sun[2,3], Minghao Yin [2,3], and Shuai Lu[2,3]

[1] College of Software, Jilin University,
Changchun 130012, China
cuishulin@email.jlu.edu.cn
[2] College of Computer Science and Technology, Jilin University,
Changchun 130012, China
jgsun@jlu.edu.cn, mhyin@nenu.edu.cn
[3] Key Laboratory of Symbolic Computation and Knowledge Engineering of
Ministry of Education, Jilin University, Changchun 130012, China

Abstract. Stochastic Shortest Path problems (SSPs), a subclass of Markov Decision Problems (MDPs), can be efficiently dealt with VI, PI, RTDP, LAO* and so on. However, in many practical problems the estimation of the probabilities is far from accurate. In this paper, we present uncertain transition probabilities as close real intervals. Also, we describe a general algorithm, called gLAO*, that can solve uncertain MDPs efficiently. We demonstrate that Buffet and Aberdeen's approach, searching for the best policy under the worst model, is a special case of our approaches. Experiments show that gLAO* inherits excellent performance of LAO* for solving uncertain MDPs.

1 Introduction

Markov Decision Problem models are based on state transition probabilities. LAO* algorithm has been proved to be one of the most efficient methods to solving MDPs without evaluating complete states [1]. However, in many practical problems the transition probabilities have to be estimated from data and the exact value may be difficult or impossible to obtain. Uncertain probabilities may have a huge impact on the solution, which is often quite sensitive to changes in the transition probabilities. This problem has long been recognized in the control community and in the reinforcement learning community.

A number of authors have addressed the issue of uncertainty in the transition matrices of MDPs. [2] and [3] have considered the transition matrix to lie in a given set. [4]has considered the transition matrix to lie in an interval, and have showed that robust RTDP can solve uncertain MDPs. [5] has introduced interval value functions as a natural extension of traditional value functions; an interval value function assigns a closed real interval to each state, representing the assertion that the value of that state falls within that interval. The approach described above has found a policy that behaves well under the worst model. But the best policy under the worst model is pessimistic, which is different from the policy under the true model in many cases.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 948 – 957, 2006.

We should give a set of available approaches, from which users can choose depending on actual problems.

We describe an algorithm gLAO* that can solve uncertain MDPs efficiently. Our approaches to solve MDPs with uncertain state transition probabilities are well applicable, and can search for the best policy in worst, best and average cases. Buffet and Aberdeen's approach is a special case of our approaches. Experiments demonstrate that gLAO* inherits excellent performance of LAO* for solving uncertain MDPs.

2 Uncertain MDPs

A uncertain MDP should be defined as a tuple[6]: $< S, s_0, G, A, c_i(a), p_{ij}(a) >$, where S is a finite state space, $s_0 \in S$ is an initial state, $G \subseteq S$ is a set of goal states, A is a finite action set and $c_i(a)$ is action cost , $p_{ij}(a)$ is the state transition probability. $p_{ij}(a)$ falls within a closed real interval $[{p_{ij}}^l(a), {p_{ij}}^u(a)]$. Let ${p_{ij}}^l(a)$ refer to the lower bounds and ${p_{ij}}^u(a)$ refer to the upper bounds.

Let f be the evaluation function that provide admissible estimates of the expected cost to reach the goal from every state i. A policy π can be represented as a stationary mapping from states to actions. The aim for solving uncertain MDP amounts to finding an optimal policy. The evaluation function value of every relevant state of π must be an interval because state transition probability is a closed real interval. Let $[f^l(i), f^u(i)]$ be the evaluation function interval of state i. Let $[Q^l(i,a), Q^u(i,a)]$ denote the interval of the "Q-value". We define the interval of the "Q-value" as follows:

$$[Q^l(i,a), Q^u(i,a)] = \{c_i(a) + \sum_{j \in S} [{p^l}_{ij}(a), p_{ij}^u(a)] \times [f^l(j), f^u(j)]\}. \quad (1)$$

Now, we consider several special orderings on closed real intervals. The different orderings on closed real intervals are defined as follows [5]:

$$([l_1, u_1] \le [l_2, u_2]) \Leftrightarrow (u_1 \le u_2 \ \ or \ \ (u_1 = u_2 \wedge l_1 \le l_2)). \quad (2)$$

$$([l_1, u_1] \le [l_2, u_2]) \Leftrightarrow (l_1 \le l_2 \ \ or \ \ (l_1 = l_2 \wedge u_1 \le u_2)). \quad (3)$$

The two orderings above are partial. [5] proves that there exists at least one optimal policy under each ordering. We also give another approach ranking the intervals,

$$([l_1, u_1] \le [l_2, u_2]) \Leftrightarrow (\frac{l_1 + u_1}{2} \le \frac{l_2 + u_2}{2}) . \quad (4)$$

Finding out the optimal policy and its evaluation function depends on how to rank the interval of the "Q-value". Using either of these orderings of intervals can yield different evaluation function and policy. We don't discuss other orderings that may also yield interesting results.

Definition 1. For any policy π and state i, we define greedy action $\pi(i)$ of π at i satisfying the following equation:

$$[Q^l(i,\pi(i)), Q^u(i,\pi(i))] = \min_{a\in A(i)}[Q^l(i,a), Q^u(i,a)]. \tag{5}$$

We here use formulas (2), (3) and (4) to compute the minimum intervals of "Q-value".

For the worst case: using formula (2),

$$\min_{a\in A(i)}[Q^l(i,a), Q^u(i,a)] \overset{(2)}{\Leftrightarrow} \min_{a\in A(i)}[Q^u(i,a)]. \tag{6}$$

For the best case: using formula (3):

$$\min_{a\in A(i)}[Q^l(i,a), Q^u(i,a)] \overset{(3)}{\Leftrightarrow} \min_{a\in A(i)}[Q^l(i,a)]. \tag{7}$$

For the average case: using formula (4):

$$\min_{a\in A(i)}[Q^l(i,a), Q^u(i,a)] \overset{(4)}{\Leftrightarrow} \min_{a\in A(i)}[\frac{Q^l(i,a)+Q^u(i,a)}{2}]. \tag{8}$$

Therefore, in order to easily find out the optimal policy, the lower bound $Q^l(i,a)$, the upper bound $Q^u(i,a)$ and average value of lower bound and upper bound instead interval of the "Q-value" would be used in the process of compute the optimal policy. Now, we give the definition of greedy action of policy π at state i under every case.

Definition 2. Depending on (6) greedy action $\pi(i)$ of π at i could be defined as follows:

$$\pi(i) := \arg\min_{a\in A(i)}[Q^u(i,a)]. \tag{9}$$

Definition 3. Depending on (7) greedy action $\pi(i)$ of π at i could be defined as follows:

$$\pi(i) := \arg\min_{a\in A(i)}[Q^l(i,a)]. \tag{10}$$

Definition 4. Depending on (8) greedy action $\pi(i)$ of π at i could be defined as follows:

$$\pi(i) := \arg\min_{a \in A(i)} \left[\frac{Q^l(i,a) + Q^u(i,a)}{2} \right]. \tag{11}$$

[5]also considers how to compute the worst model. Let $R = \{ j_k \mid k = 1,...,n \}$ be the reachable state set for a given state i and an action a. Firstly, we sort reachable states in decreasing order of the value: $f(j_1) \geq f(j_2) \geq ... \geq f(j_n)$. Then we give the much probability mass to the states early in the ordering. Computing $Q^u(i,a)$ is the process finding the worst model. To find the best model, similarly, the idea is to sort the possible states into increasing order according to their $f(j)$ value. Then, the best probability distribution is the one giving the highest probability to the first state j_1, then to j_2, and so on up to j_n. Computing $Q^l(i,a)$ is the process finding the best model. Obviously, Buffet and Aberdeen's approach is a special case of our approaches. Therefore it is straight forward to have the following lemmas.

Lemma 1. There exists a greedy policy on the interval of the "Q-value", which is the same as the one under the worst model according to (2).

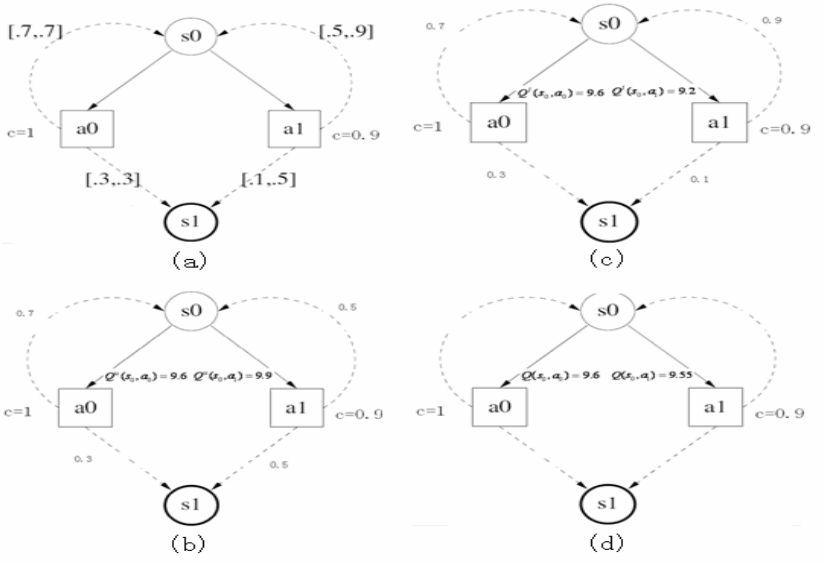


Fig. 1. (a) A simply sample of Uncertain MDPs used in [4]; (b) The transition probability distributions and "Q-value" depending on (9); (c) The transition probability distributions and "Q-value" depending on (10); (d)The "Q-value" depending on (11)

Lemma 2. There exists a greedy policy on the interval of the "Q-value", which is the same as the one under the best model according to (3).

We consider an Uncertain MDP sample used in [4]. We assume that the current evaluation function values of s_0 and s_1 are $f(s_0)=8$ and $f(s_1)=10$. The "Q-value" in different cases are different, and hence the greedy actions at s_0 are different. Figure 1 shows the transition probability distributions and "Q-value" in different cases. Depending on (9), $Q^u(s_0,a_0)=9.6$, $Q^u(s_0,a_1)=9.9$, and the greedy action is a_0; depending on (10), $Q^l(s_0,a_0)=9.6$, $Q^l(s_0,a_1)=9.2$, and the greedy action is a_1; depending on (11), $Q(s_0,a_0)=9.6$, $Q(s_0,a_1)=9.55$, and the greedy action is also a_1.

3 Algorithms That Solve Uncertain MDPs

3.1 Dynamic Programming Algorithms

Given the complete and accurate state-transition probabilities $p_{ij}(a)$, and the action costs $c_i(a)$ for all states i and actions $a \in A(i)$, it is possible to find an optimal policy via a dynamic programming algorithm for MDPs. There are two dynamic programming algorithms: value iteration and policy iteration [7]. Both value iteration and policy iteration repeat an improvement step that updates the evaluation function for all states.

We now describe two related dynamic programming algorithms for Uncertain MDPs: gVI(general Value Iteration) and gPI(general Policy Iteration). Let Worst VI, Best VI and Average VI denote the algorithms gVI to compute the optimal policy in worst, best, and average cases. Let Worst PI, Best PI and Average PI denote the algorithms gPI to compute the optimal policy worst, best and average cases. gVI and gPI are summarized in figure 2 and figure 3, respectively.

3.2 gLAO*

When an Uncertain MDP is formalized as a graph-search problem, each state of the graph corresponds to a problem state and each arc corresponds to an action causing a transition from one state to another. [8] views an AND/OR graph as a hypergraph that has hyperarcs or k-connectors, connecting a node to a set of k successor nodes. K-connector is interpreted as the transformation of a problem into k sub problems.

Figure 4 outlines the algorithm gLAO* for finding a minimal-cost solution graph in a AND/OR graph with loops. The version of gLAO* we have outlined follows the LAO* described by [1]. gLAO* differs from the original LAO* in that: gVI or gPI is used to compute state costs in the dynamic programming step.

Algorithm 1. gVI

1. Start with an initial evaluation function f_0 and set $\varepsilon > 0$.
2. Repeat until the error bound of the evaluation function is less than ε for each state $i \in S$:

   %the worst case: $f'(i) = \min_{a \in A(i)} [Q^u(i,a)]$;

   %the best case: $f'(i) = \min_{a \in A(i)} [Q^l(i,a)]$;

   %the average case: $f'(i) = \min_{a \in A(i)} \left[\frac{Q^l(i,a) + Q^u(i,a)}{2} \right]$.

3. For each state $i \in S$,

   %the worst case: using formula (9);

   %the best case: using formula (10);

   %the average case: using formula (11).

Fig. 2. gVI: A dynamic programming algorithm for Uncertain MDPs

Algorithm 2. gPI

1. Start with an initial evaluation function f_0 and an initial policy π_0.
2. Repeat until the policy does not change for each state $i \in S$:

   %the worst case:

   i. Policy evaluation: $f^\pi(i) = [Q^u(i, \pi(i))]$,

   ii. Policy improvement: using formula (9);

   %the best case:

   i. Policy evaluation: $f^\pi(i) = [Q^l(i, \pi(i))]$,

   ii. Policy improvement: using formula (10);

   %the average case:

   i. Policy evaluation:

$$f^\pi(i) = \left[\frac{Q^l(i,\pi(i)) + Q^u(i,\pi(i))}{2} \right],$$

   ii. Policy improvement: using formula (11).

3. Return an optimal policy π^*.

Fig. 3. gPI: A dynamic programming algorithm for Uncertain MDPs

When gPI is used to compute exact state costs in the dynamic-programming step, the solution to which gLAO*converges is optimal. When state costs are updated approximately using gVI in dynamic-programming step, the solution to which gLAO* converges may not be optimal. In this case, an additional convergence test can be used to determine whether the solution is $\varepsilon -$ optimal, for any ε.

Let Worst LAO*, Best LAO*, and Average LAO* denote the algorithm gLAO*, which perform gVI (or gPI) in three different cases to compute the optimal policy.

Algorithm 3. gLAO*

1. The explicit graph G' initially consists of the start state s_0.
2. While the best solution graph has some non-terminal tip state:
   (a) Expand best partial solution: Expand some non-terminal tip state n of the best partial solution graph and add any new successor states to G'. For each new state i added to G' by expanding n, if i is a goal state then $f(i) := 0$; else $f(i) := h(i)$.
   (b) Update state costs and mark best actions:
   i. Create a set Z that contains the expanded state and all of its ancestors in the explicit graph along marked action arcs. (I.e., only include ancestor states from which the expanded state can be reached by following the current best solution.)
   ii. Perform dynamic programming (gVI or gPI) on the states in set Z to update state costs and determine the best action for each state.
3. Convergence test: If gPI was used in step 2(b) ii, go to step 4. Else perform gVI on the states in the best solution graph. Continue until one of the following two conditions is met. (i) If the error bound falls below ε, go to step 4. (ii) If the best current solution graph changes so that it has an unexpanded tip state, go to step2.
4. Return an optimal (or $\varepsilon -$ optimal) solution graph.

Fig. 4. gLAO*: a heuristic search algorithm for Uncertain MDPs, which can find an optimal solution graph without evaluating the entire state space

4 Experiments

We use here the race track problem as defined in [9]. A race track of any shape includes a start state at one end and a finish line at the other consisting of designated squares. Each square within the boundary of the track is a possible location of the car. Beginning at the start state, the car attempts to move along the track to the finish line. The state of the system can be determined by the car's location and its speeds in the horizontal and vertical directions. There is a set of admissible nine actions for each state. With probability 0.9, the controller's intended action is reliably executed, and with probability 0.1, the system ignores the controller's action. That is, 0.1 is the

Table 1. Summary of performance for Worst LAO*, Average LAO* and Best LAO*

		running time	number of two steps	update numbers	numbers of relevant states
	WLao*	20.031	68	374367	21785
R	**ALao***	21.203	68	393783	21807
	BLao*	20.515	68	380449	21805
	WLao*	4.625	38	85307	9259
L	**ALao***	4.985	39	89372	9254
	BLao*	4.812	39	88844	9263
	WLao*	0.109	20	2702	408
R-1	**ALao***	0.125	21	2796	408
	BLao*	0.110	21	2796	408
	WLao*	0.468	30	10466	1254
R-2	**ALao***	0.500	31	10678	1255
	BLao*	0.468	30	10323	1254
	WLao*	2.875	38	56810	5734
R-3	**ALao***	2.781	37	54341	5730
	BLao*	2.859	38	56142	5730
	WLao*	23.890	53	389266	32160
R-4	**ALao***	25.844	53	417452	32197
	BLao*	24.843	53	409479	32202
	WLao*	0.750	31	15175	2473
S-1	**ALao***	0.718	32	14800	2474
	BLao*	0.625	31	14056	2471
	WLao*	3.406	33	55922	8841
S-2	**ALao***	3.406	31	56418	8836
	BLao*	3.047	30	52996	8839
	WLao*	28.578	40	295524	41872
S-3	**ALao***	28.687	40	341923	41872
	BLao*	28.578	39	329855	41880

probability that the actual accelerations or deceleration at a move are zero. If the car hits the track boundary, we move it back to a random position on the starting line, reduce its velocity to zero, and continue the trial. The objective is to learn to control the car so that it crosses the finish line in as few moves as possible.

We applied Worst LAO*, Best LAO*, and Average LAO* to race track problems. We set the initial costs of all the states to zero and $\varepsilon = 10^{-3}$. And the controller's intended actions were reliably executed with uncertain probability [0.85, 0.95].

In the figure 5 we compare the optimal solutions generated by the three different algorithms for a given map L, which are the paths the car would follow from the start

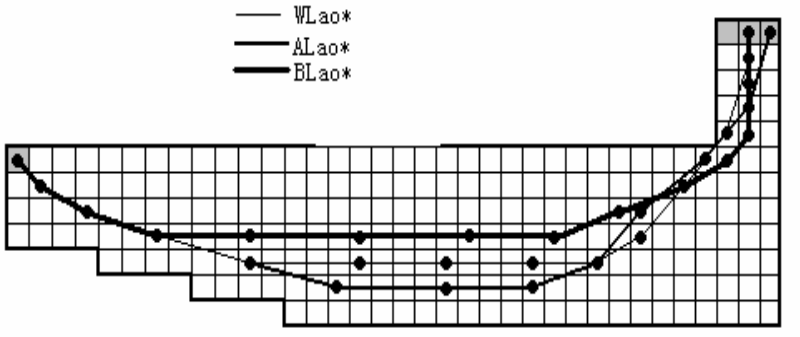


Fig. 5. This figure illustrates the optimal paths derived by Best LAO*, Worst LAO* and Average LAO* for map L

state if all sources of randomness in the problems were turned off. Noting that Worst LAO*, Best LAO*, and Average LAO* take very different paths to the goal. Table1 shows performance on some race track maps for Best LAO*, Worst LAO* and Average LAO*. Through the experimental results we can conclude that:

i. Like LAO*, gLAO* is a heuristic search algorithm, it can solve uncertain MDPs without evaluating the entire state space.
ii. Worst LAO*, Best LAO*, and Average LAO* can solve effectively uncertain MDPs, we can choose different approaches depending on different situations.

5 Conclusion

We have proposed a general algorithm gLAO* that could deal with uncertain MDPs. The approaches adopted in this algorithm consider three different cases: the best case, the worst case and the average case. Different approaches would be chosen depending on different situations. gLAO* inherits the merits of LAO*, it could find out an optimal solution in MDPs with uncertain transition probabilities.

Acknowledgment

We thank Blai Bonet for providing the source codes of some algorithms. This work is fully supported by the National Natural Science Foundation of China under Grant No. 60473003; the National Research Foundation for the Doctoral Program of Higher Education of China under Grant No. 20050183065; the Outstanding Youth Foundation of Jilin Province of China under Grant No. 20030107; the Science Foundation for Yong Teachers of Northeast Normal University Grant No. 20051001.

References

1. Eric A. Hansen, and Shlomo Ziberstein. LAO*: A heuristic search algorithm that finds solutions with loops. Artificial Intelligence 129(2001)35-62
2. J.A. Bagnell, A. Y. Ng, and J. Schneider. Solving uncertain markov decision problems. Technical Report CMU-RI-TR-01-25, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, August 2001.
3. Uresh Kalyanasun daram, Edwin K.P. Chong and Ness B. Shroff. Markov Decision Processes with Uncertain Transition Rates: Sensitivity and Robust Control. Proceedings of the 41st IEEE, Conference on Devision and Control, Las Vegas, Nevada USA, December 2002
4. O. Buffet and D. Aberdeen. Robust planning with (l)rtdp. In Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI'05), 2005.
5. R. Givan, S. Leach, and T. Dean. Bounded parameter markov decision processes. Artificial Intelligence, 122(1-2):71–109, 2000.
6. D.P. Bertsekas and J.N. Tsitsiklis. Neurodynamic Programming. Athena Scientific, 1996.
7. D. Bertsekas, Dynamic Programming and Optimal Control, Athena Scientific, Belmont, MA, 1995.
8. A. Martelli, U. Montanari, Optimizing decision trees through heuristically guided search. Comm. ACM 21 (12) (1978) 1025-1039.
9. A.G. Barto, S. Bradtke, and S. Singh. Learning to act using real time dynamic programming. Artificial Intelligence, 72, 1995.

Autonomous Navigation Based on the Velocity Space Method in Dynamic Environments

Shi Chao-xia[1], Hong Bing-rong[1], Wang Yan-qing[2], and Piao Song-hao[1]

[1] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang,150001, China
stonexia@sohu.com
[2] School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, Heilongjiang, 150080, China
wyq0325@126.com

Abstract. We present a new local obstacle avoidance approach for mobile robots in partially known environments on the basis of the curvature-velocity method (CVM), the lane-curvature method (LCM) and the beam-curvature method (BCM). Not only does this method inherit the advantages from both BCM and LCM, but also it combines the prediction model of collision with BCM perfectly so that the so-called prediction based BCM (PBCM) comes into being and can be used to avoid moving obstacles in dynamic environments.

1 Introduction

There are important benefits in using autonomous mobile robots to perform different tasks. For example, robots may be used to prospect the unknown in hazardous environments such as volcanoes[1], deep underwater sites[2], or outer space[3], where human operators could be at risk. In conventional environments the use of mobile robots to continuously perform fetching and carrying jobs offers significant economic advantages through the release of skilled human workforce[4].

Some researchers use purely reactive methods, such as the Artificial Potential Fields methods (APF)[5,6] and the Vector Field Histogram method (VFH)[7] to guide the robot exploring in unknown environments. Though the two methods produce smoother travel and can handle both narrow and wide openings, they can't account for the fact that when the robot turns it typically move along arcs, rather than in straight lines. Velocity space approaches, including CVM[8], LCM[9] and BCM[10], formulate the local obstacle avoidance problem as one of constrained optimization in the velocity space and can take vehicle dynamics into account by choosing angular velocity along with linear velocity. Though BCM generates smoother and faster trajectory than other methods above, it may fail to navigate in dynamic environments due to the fact that BCM takes only the obstacles' position rather than velocity into account. So we combine BCM with a prediction model of collision so that the robot has the capability of avoiding the dynamic obstacle in partially known environments.

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 958 – 961, 2006.

2 Improvements on BCM

We divide all the beams into two types: the free beam and the block beam. The block beam is the beam associates with an obstacle while the free beam is vice versa. Enlightened by LCM, we take the beam's width into account so that the beam can be denoted as a quadruple, $B(\rho_{i1}, \rho_{i2}, d_i, w_i)$, where ρ_{i1} and ρ_{i2} are the orientations of corresponding sensors. The width of the *i*-th beam is calculated as follows:

$$w_i = \begin{cases} 2d_i \tan(\dfrac{\rho_{i2} - \rho_{i1}}{2}) & block \quad beams \\ \sqrt{(x_{i-1,m_{i-1}} - x_{i+1,1})^2 + (y_{i-1,m_{i-1}} - y_{i+1,1})^2} & free \quad beams \end{cases} \quad i = 1,2,\cdots,n \ . \tag{1}$$

Where n is the total amount of the beams; $x_{i-1,mi-1}$ is the x-coordinate of the extremely right point of the (i-1)-th obstacle; $y_{i+1,1}$ is the y-coordinate of the extremely left point of the (i+1)-th obstacle.

After calculating all the beams in robot's local reference frame, we select the best beam to guide the robot by maximizing an objective function.

$$f(\rho_1, \rho_2, d, w) = \alpha\left(\frac{d\cos(|\varepsilon|)}{d_{\max}}\right) + \beta\left(\frac{w}{w_{\max}}\right) - \gamma\left|\frac{\varepsilon}{\pi}\right| \ . \tag{2}$$

Where ε, having been defined by Eqn.9 in literature[10], is the smallest angular error between the goal direction and the analyzed beam; α, β and γ are weight constants adjusted by experimentation ($\alpha + \beta + \gamma = 1$).

3 The Prediction Model of Collision

We use the extended Kalman filter to calculate the obstacle's velocity and eliminate the uncertainty resulted from the disturbance of system noise simultaneously.

The state equation and measurement equation are denoted as follows:

$$\begin{cases} X(k+1) = f(X(k), u(k)) + W(k) \\ Z(k+1) = H(k+1)X(k+1) + V(k+1) \end{cases} . \tag{3}$$

Where $X(k)=[\ x_k \quad y_k \quad \alpha_k \quad v_k \quad \omega_k\ ]^T$ is the state vector consisting of the obstacle's position, orientation and velocities in robot's local reference frame at time step k. $u(k)=[\ D_k, \Delta\alpha_k\ ]$ is the input vector in which D_k is the robot's driving distance and $\Delta\alpha_k$ is the robot's rotation angle in k-th time interval; $W(k)$ and $V(k+1)$ are the zero-mean, white, Gaussian process noise with known covariance.

$$H(k+1) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} . \tag{4}$$

$$f(X(k),u(k))=\begin{bmatrix}(x_k+v_kT\cos(\omega_kT))\cos(\Delta\alpha_k)-(y_k+v_kT\sin(\omega_kT))\sin(\Delta\alpha_k)\\(x_k+v_kT\cos(\omega_kT))\sin(\Delta\alpha_k)+(y_k+v_kT\sin(\omega_kT))\cos(\Delta\alpha_k)+D_k\\\alpha_k+\omega_kT-\Delta\alpha_k\\v_k\\\omega_k\end{bmatrix} \tag{5}$$

T is the sampling interval; $H(k+1)$ and $Z(k+1)$ are 5×2 const matrix and 2×1 measurement vector respectively.

Although some emergencies, such as sudden turn, acceleration or deceleration, take place sometimes, in most cases, the obstacle's velocity can be regarded as a uniform speed. Eqn.6-8 shows us the theory of trajectory prediction.

$$\theta_o(k)=\arctan(\frac{y_o(k)-y(k-1)}{x_o(k)-x(k-1)})+(\omega_r(k)-\omega_o(k))t(k)\ . \tag{6}$$

$$\begin{bmatrix}x_o(k+1)\\y_o(k+1)\end{bmatrix}=\begin{bmatrix}x(k)\\y(k)\end{bmatrix}+\begin{bmatrix}v_o(k)t(k)/\sqrt{1+\tan^2(\theta_o(k))}\\v_o(k)t(k)\tan(\theta_o(k))/\sqrt{1+\tan^2(\theta_o(k))}\end{bmatrix}. \tag{7}$$

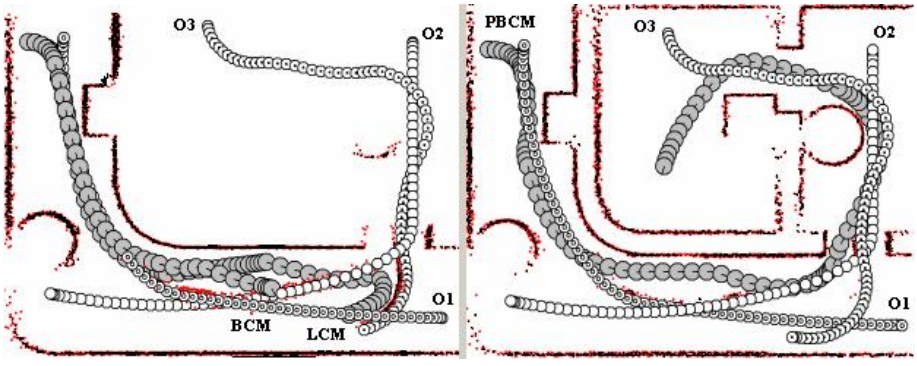
$$\begin{bmatrix}x(k+1)\\y(k+1)\end{bmatrix}=\begin{bmatrix}\cos(\omega_r(k+1)t(k)) & -\sin(\omega_r(k+1)t(k))\\\sin(\omega_r(k+1)t(k)) & \cos(\omega_r(k+1)t(k))\end{bmatrix}\begin{bmatrix}x_o(k+1)\\y_o(k+1)\end{bmatrix}+\begin{bmatrix}0\\-v_r(k)t(k)\end{bmatrix} \tag{8}$$

4 Experiments and Results

Autonomous navigations have been implemented in a dynamic environment with real robot. As is shown in Fig.1a, BCM collides with the obstacle O_2 after passing by the first one. LCM, however, avoids both obstacle O_1 and obstacle O_2 luckily and meets obstacle O_3 in front of the door, where there exists no applicable lane to go through so that it has to turn round for the new path. PBCM, as is shown in Fig.1b, exhibits considerable advantages in avoiding moving obstacles.

To evaluate the efficiency of PBCM, LCM and BCM, we obtain the average time spent in different scenes. LCM spends more time than that of BCM and PBCM in static environment with or without obstacles. BCM and PBCM take almost the same time in capacious environment. PBCM achieve 28 in 30 total dynamic experiments, the only 2 failures take place in narrow doorway when obstacle O_3 moves out suddenly. In such condition, when the robot recognizes the moving obstacle, the obstacle has been too near to avoid. While BCM and LCM achieve 11 and 16 respectively in all cycle, which indicates the validity of our approach.

The following conclusions can be drawn from the experiments above. Firstly, the combination of the prediction model and the improved BCM endows the robot the capability of dynamic obstacles avoidance. Secondly, PBCM, as a reactive navigational method, often fails to overcome local minima problems. Combining PBCM with a global environmental map may be a feasible method to resolve these problems.



a) The trajectory of BCM and LCM b) The trajectory of PBCM

Fig. 1. Experimental result in dynamic environment

References

1. Daniele Caltabiano, Giovanni Muscato, Francesco Russo, Localization and self-calibration of a robot for volcano exploration, In Proc. IEEE Intl. Conference on Robotics & Automation, New Orleans, LA, April (2004).
2. Hiroshi Kawano, Tamaki Ura, Motion planning algorithm for non-holonomic autonomous underwater vehicle in disturbance using reinforcement learning and teaching method, In Proc. IEEE Intl. Conference on Robotics & Automation, Washington DC, May (2002).
3. Karl Iagnemma, Shinwoo Kang, Hassan Shibly, and Steven Dubowsky, Online Terrain Parameter Estimation for Wheeled Mobile Robots With Application to Planetary Rovers, IEEE Transaction on robotics, Vol.20, No.5, October, (2004)
4. Angelos Amditis, Fusion of infrared vision and radar for estimating the lateral dynamics of obstacles, Information Fusion 6 (2005) 129–141
5. S.S. Ge and Y.J. Cui, Dynamic motion planning for mobile robots using potential field method, Autonomous Robots, (2002) 207-221.
6. O. Khatib, Real-time obstacle avoidance for manipulators and mobile robots. In Proc. IEEE Intl. Conference on Robotics and Automation, St. Louis, MO, March (1985) 500–505.
7. J. Borenstein, Y. Koren, The vector field histogram—fast obstacle avoidance for mobile robots, IEEE Journal of Robotics and Automation 7 (3) (1991) 278–288.
8. R.G. Simmons, The curvature velocity method for local obstacle avoidance, In Proc. IEEE Intl. Conference on Robotics & Automation, vol. 4, (1996) 2275–2282.
9. N.Y. Ko, R.G. Simmons, The lane-curvature method for local obstacle avoidance, In Proc. of the Intl. Conference on Intelligent Robotics and Systems, Canada, October, (1998).
10. J.L. Fernandez, Improving Collision Avoidance for Mobile Robots in Partially Known Environments: the Beam Curvature Method, Robotics and Autonomous Systems 46 (2004)

Intentional Agency Framework Based on Cognitive Concepts to Realize Adaptive System Management

Yu Fu[1], Junyi Shen[2], and Zhonghui Feng[2]

Department of Computer Science and Technology, Xi'an Jiaotong University
Xi'an 710049, P.R. China
[1]fuyu_eternal@people.com.cn
[2]{jyshen, fzh}@mail.xjtu.edu.cn

Abstract. Management software is required to be adaptive and scalable since the managed objects become more and more dynamic and complex. In this paper, we introduce cognitive concepts to deal with the complexity in developing management software, and an interactive intelligent agency framework, called iSMAcy (intelligent System Management Agency), is developed as an integrated solution to realize adaptive system management. The cognitive concepts are applied at two levels. At micro level, the intentional agent kernel of functional agent use mental states such as beliefs, goals and intentions to realize integration of goal-directed behavior with situated reactive action. At macro level, high-level cognitive semantics such as goals, requests and commitments are interchanged among agents. An example of applying iSMAcy system on a simulated network management environment is described, and the efficiency of the intentional agent kernel is analyzed to ensure the reactivity of iSMAcy system.

1 Introduction

As the managed objects become more and more dynamic, complex and heterogeneous, we need the management software to be more adaptive and scalable. Conventional development techniques are difficult and expensive to build and maintain such systems. The challenges roughly come from two levels: the micro level concerns on how to capture the complexity in ensuring the autonomy of local controller, and the macro level concerns on how to capture the complexity of interaction among different authorities. In this paper, we adopt an intentional multi-agent approach to treat with the complexity in both of these two levels. The intelligent agency framework we built, called iSMAcy (intelligent System Management Agency), are modeled from intentional stance [10, 12], which has been used as a powerful abstract tool to describe, explain, and predict the systems behavior by applying cognitive mental status such as beliefs, goals, and intentions.

The remainder of the paper is organized as follows. We give the macro system components of iSMAcy in Section 2. The micro architecture of intentional agent kernel to realize local autonomy is described in Section 3. At Section 4, we illustrate how the distributed and cooperated management paradigm is realized. A case study of applying iSMAcy system on a simulated testbed to realize network resource

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 962 – 971, 2006.

management is reported in Section 5. In Section 6, a detailed analysis on reactivity of the intentional agent kernel is given. The related work is discussed at Section 7 and finally the paper is concluded at Section 8.

2 The Agency Framework Architecture

iSMAcy is a FIPA (Foundation for Intelligent Physical Agents) [2, 3] compliant agency framework, composed by server agency node connected with high-speed networks. As depicted in Fig. 1, an iSMAcy agency node includes: (1) an agency platform (AP) that provides physical infrastructure to deploy agents, and (2) an application agent container that collects functional agents for management tasks. The agency platform composes an Agent-Management-System (AMS), a Directory Facilitator (DF), and a Message Transport Service (MTS). The AMS is responsible to control the agent lifecycle from agent creation to termination. The DF helps agents to find appropriate partners for a particular conversation. The MTS, which contains the Agent Communication Channel (ACC) and Message Transfer Protocols (MTPs), is used to deliver messages between agents inner or inter APs. In application agent container, the diagnosis agent is responsible for receiving alarms and status from the managed objects, diagnosing the problems, then sending the diagnosed results to the control agent. The later will carry out corrective actions to solve the problems. iSMAcy supports distributed management paradigm: the control agent of each iSMAcy node can send messages and process conversations with other control agent residing on remote agency node.

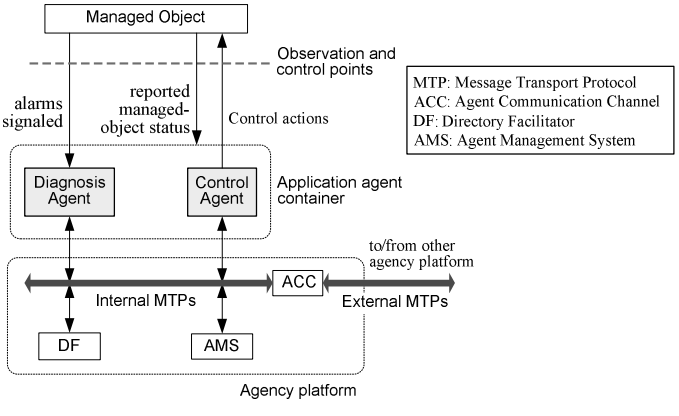


Fig. 1. Architecture of an iSMAcy node

3 Local Autonomy

Both of the diagnose agent and control agent are built from an Intentional Agent (IA) kernel, which is modeled from intentional stance [10, 12]. Its basic components include the data structures representing its mental attitudes and a set of control processes explaining the interrelationship between these attitudes.

3.1 Data Structures of Mental Attitudes

Beliefs, capacities, goals, and intentions are chose to represent the mental attitudes. Beliefs present the world model of the agent perceived from its environment. Its contents are organized as a list of facts with the following format:

```
World_Model:
      FACT <relation_id_1> <parameter_id:type:value>+
      … …
      FACT <relation_id_n> <parameter_id:type:value>+
```

Knowledge about how to accomplish a given goal or react to certain situation is encoded in the plan structure and is placed in the capacities library. Each plan comprises a plan head and a plan body. The plan head defines under what circumstances the plan body is applicable. Its contents including: a unique name, a trigger condition field, a precondition field, and a plan utility field. The basic elements of plan body are primitive actions and sub-goals, which can be combined with language control structures like if-then-else, while-do iteration, and sequential/disjunctive execution blocks. An example of plan structure used in the experiment of Section 5 is depicted at Fig. 2.

```
1   PLAN: {
2   NAME: [Control_link_degrade]
3   INVOCATION CONDITION:
4    GOAL: [Control_link_degrade $Congested_Link $Env_Diaged_problem]
5   PRECONDITION: [Null]
6   CONTEXT: [TEST (<= Get_Congestion_Level(self_SubNetArea) #Middle)]
7   UTILITY: [Normal]
8   BODY:
9    Let $control_path as the lowest priority one that has the $Congested_Link_Pair
       as first_offered_route
10   ASSIGN $next_route Get_next_route(route_table($control_path))
11   IF (TEST (NOT $next_route NULL))
12        Apply_SK_command ($switch $Congested_Link)
13   ELSE {
14        Get ($idle_route) that is out of route_table($switch) with shortest length
          or maximum idle capapcity
15        IF (TEST (NOT $idle_route NULL))
16           Apply_IRR_command ($switch $idle_route $Congested_Link_Pair)
17        ELSE {
18           Send_interPlatform_Message
             GetControlAgent(GetSourcePlatform($control_path)) self_platform
             #Env_Diaged_problem
19           … … …
20           apply protective controls on self_SubNetArea
21           }}
22  }
```

Fig. 2. A plan to control degraded links

The goals set represent the motivation component of the agent. We distinguish "top-level goals" and "subgoals". The former represent the highest-level goals that are specified on the initiate times when the agent start up; the later are goals that the agent creates during plan execution. The BNF grammar of a goal item can be presented as follows:

```
goal-value ::= <goal-type><relation-identify><parameter>*<goal-
   utility>
goal_type ::= <achieve>|<perform>|<maintain>
```

The intention structure, which is made up by several concurrently running intention stacks, represents current actions the agent has committed to perform. Each intention stack composes of a top-level goal that has been pursued and all its subgoals and instantiated plans that have been extended to achieve these goals.

3.2 Control Processes Manipulating the Mental Attitudes

The control processes manipulating the mental states include Observer and Message receiver, APL (Applicable Plan List) Generator, APL Filter, Mental-State Updater and Intention Executor. The execution model of the IA kernel is depicted in Fig. 3. At the beginning of each cycle, the Observer and Message-receiver gets new perceptions and ACL (agent communication language) messages, resulting new events are produced in the event queue. The APL Generator matches each event with the trigger condition of plans in the capacities library, and the successful one is initiated and placed into the APL structure. The APL Filter decides whether the plans compatible with current intentions or the plans corresponding to new external events will survive. The former let agent work in a goal-directed manner and the later in a reaction manner. Plans that passed the Filter will be inserted into the corresponding position of intention structure by the Mental-state Updater, which also responds to process the internal events. For example, a "Goal-Success" event will pop the corresponding end-node (with all of its subgoal nodes) in the intention structure. Finally, the Intention Executor gets the leaf-nodes of the intention structure with highest utility to form a frame of execution block, which contains limited number of primitive actions and subgoals, and then execute the actions with the execution block step by step, resulting actions to be taken or goals to be posted to the event queue.

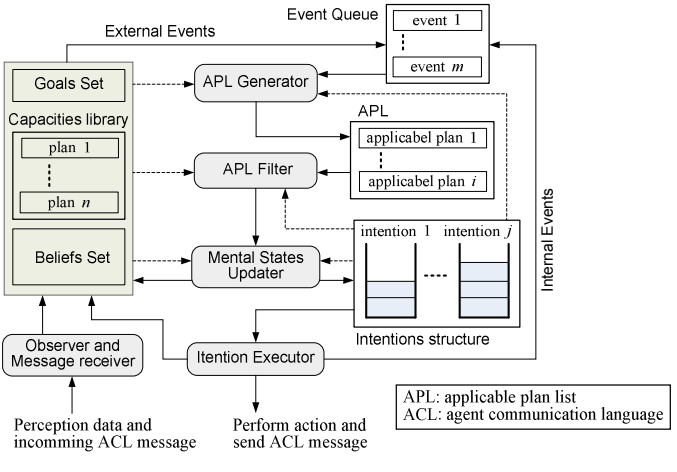


Fig. 3. Execution model of Intentional Agent kernel

4 Distributed and Cooperated Management Paradigm

iSMAcy provide infrastructures to support agent communicating and exchanging information with FIPA Agent Communication Language (ACL). For each agent there is a uniform messaging interface used for message delivery and dispatching. In complex management environments, agents usually need to interact with others with a conversation, that is, at any point of the conversation, after some messages have reached, certain other message sequences are expected to follow. iSMAcy also support agent to engage conversation through interaction protocols. It provides predefined plans to realize FIPA interaction protocols, such as FIPA-Request, FIPA-Query, FIPA-Propose, and FIPA-Contract-Net.

Different with traditional middleware and distributed objects technologies, iSMAcy provide plentiful semantic support for the agent interaction. For a network management example of Section 5, an agency node A may hope another agency node B to achieve the goal of "accepting 75% overflows from A". The receiver has local problem-solving capabilities—itself decide what actions to take to achieve these goals. Even more, if the goal conflict with its existing goals, the controller may choose not to accept this kind of request.

5 Experiment and Results

A testbed on network resource management has been build to demonstrate how to apply the iSMAcy prototype to a practical domain. The simulation system is deployed on four computer nodes (hardware: x86 Family15 Model2 Stepping4 GenuineIntel ~1992MHz, 512M RAM; platform: Java HotSpot™ Client VM 1.4.2_06-b03), connected with a 100Mb/s Ethernet LAN. Three simulation modules—a Traffic Simulator, a Subnetwork Area Simulator, and an iSMAcy node—are running at each computer node. The topology of the testbed is illustrated in Fig. 4. The whole

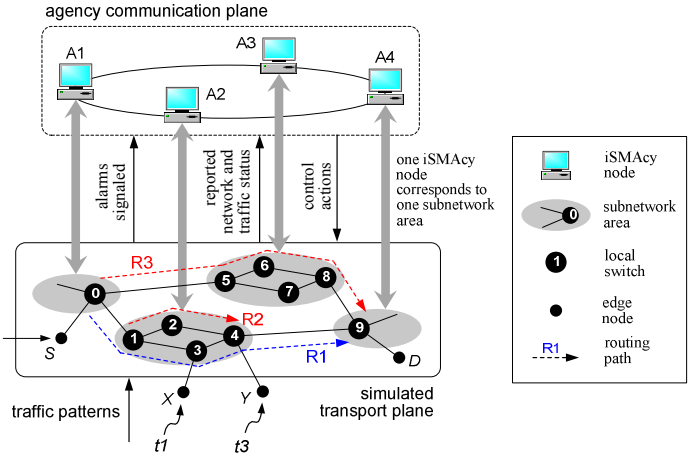
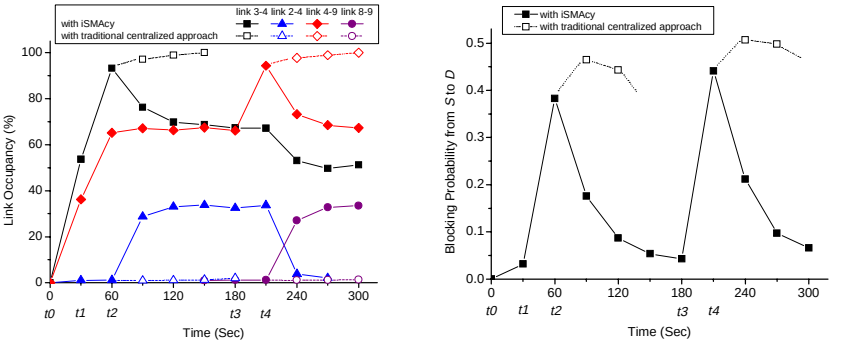


Fig. 4. The simulated network management example

managed networks are divided into four regions, each of which is managed by an iSMAcy node (labeled as agency *A*1-*A*4). For each agency node, the knowledge about the simulated network and offered traffic are maintained in the beliefs of functional agents, including the physical and logical topology of network, the node and link capacity, etc. The knowledge of how to diagnose the network problems, and how to take corrective actions to rectify problems, are encoded into the capacities library of diagnosis agent and control agent respectively. An example of plan used by the control agent to deal with a degraded link is described in Fig. 2. The strategy is to try expansive controls first, if alternative path can not be found within its local area, the control agent will send messages to other agency node request for cooperation.



(a) Impact on link occupancy (b) Impact on blocking probability

Fig. 5. Effects of applying iSMAcy system

Our tests are designed to compare the management behavior of iSMAcy with the traditional centralized approach. The later cannot provide adaptive control in a finer granularity because the status of managed network elements and the control parameters are collected and updated centrally in a fixed period (here it is set as a cycle of 300 seconds). In the test scenario, the trunk group size of links inter subnetwork are all set to be 270 channels (one call per channel), and that of all links within subnetwork are set to be 180 channels. Traffic C_{SD} from edge node *S* to *D* is added at time t_0, C_{XD} from *X* to *D* is added at time t_1, and C_{YD} from *Y* to *D* is added at time t_3. All calls in the traffic use the same negative exponential distributions for the inter-arrival time and duration.

Originally, traffics of C_{SD} are routed through path *R*1 with a normal state. At time t_1 the offered traffics of C_{XD} are injected into switch 3 and the link pair 3-4 begins to congest. With traditional method, the system only read input data and re-compute control parameters every 300 seconds, thus this fluctuation is not noticed and no action is taken. As illustrated in Fig. 5, the blocking probability from *S* to *D* will climb up nearly to 50% after a few seconds. However, when iSMAcy system is applied, the diagnose agent of agency-*A*2 detect this congestion and then the control agent begin to carry out rectify actions after time t_2—about 67% traffics are rerouted

from link1-3 to link 1-2 using the path $R2$. As a result, the utilization of congested link 3-4 decreased to 70% and the blocking rate from X to D fell below 10%.

At time t_3 the offered traffics of C_{YD} are injected into switch 4, which makes link4-9 to be congested. In centralized manner, this fine-grained time scale fluctuation is not attended either. When iSMAcy system is applied, the agency node $A2$ finds that it can not solve this problem alone within its local subnetwork. It uses a "request" protocol to send a message showed in Fig. 6 to agency node $A1$, which propose the receiver to re-route at lest 80% traffics out of the congest area. By receiving this message, agency $A1$ begins to reroute all of traffics of C_{SD} to path $R3$ after time t_4. As a result, after about 60 seconds, the utilization of the congested link 4-9 decreased and the blocking rate returned to the normal level (below 10%) again.

```
(request
   :sender  (Control_Agent@iSMAcy_A2: 1089/ACC)
   :receiver(Control_Agent@iSMAcy_A1: 1099/ACC)
   :ontology      iSMAcy_NM
   :language      FIPA-SL0
   :protocol      fipa-request
   :content
       ((action Control_Agent @iSMAcy_A1: 1099/ACC
          (re-route-area
             : Operate-path 0 9
             : Congest-area iSMAcy_A2
             : Affected-volume > 80%)))
)
```

Fig. 6. FIPA ACL message sent by iSMAcy-A2

6 The Efficiency Analysis of Agent Kernel

In determining applicable plan list, the system does not perform any general deduction, except the unification and Boolean expression evaluation when deciding what action to take. Here, we use plan library and interleave planning with execution. Both beliefs and goals are matched directly with invocation conditions by using unification only. This allows appropriate plans to be selected very quickly and guarantees reactivity.

As illustrated in Fig. 7, one basic control loop of the IA kernel roughly include 5 phases: (1) get new external events, suppose the maximum duration time is t_{OB}; (2) option generation, suppose the maximum duration time is t_{GP}; (3) filter options, suppose the maximum duration time is t_S; (4) update the mental states, suppose the maximum duration time is t_M; and (5) schedule and execution of one primitive action in the intention structure, suppose the maximum duration time is t_{PRI}. Assume at time T_0, E is signaled with the highest priority, and the first action reacting to this event begin to execute at time T_R, then the reaction time of event E is $R_E = T_R - T_0$. The maximum of R_E occurs when event is signaled just at the time the getting new external event phase had completed. We have $\max(R_E) = t_{OB} + 2(t_{GP} + t_S + t_M) + t_{PRI}$. Let n the upper bound of the number of events occur within one unit time, the control loop takes at most p time to select the applicable plans for each event, and the maximum number of events occur during the reaction interval is y. We have $y = n(R_E + t_{OB})$ and

$2t_{GP} = yp$, then we get $\max(R_E) = [(1+np)\ t_{OB} + 2(\ t_S + t_M) + t_{PRI}] \ / \ (1-np)$. R_E is bounded in time if t_{PRI}, t_M, t_S, t_{OB}, and t_{GP} are bounded respectively.

t_{PRI} is bounded since the number of primitive actions in a execution block is limited. t_M is bounded since the number of items in the belief sets and intention structure is limited. t_S is bounded since the number of items in the APL is limited. t_{OB} is bounded by the communication mechanism of interface between agent and operating system. As the number of events occurred within limited time scales and the number of plans in the capacities library are both limited, t_{GP} is guaranteed to be bounded. Table 1 shows the run-time statistic values of agency node A2 measured in the network management example described above. The APL generation time roughly corresponds to $t_{OB} + t_{GP} + t_S$. The long maximum plan execution time of control-agent comes from the inter-node communication latency when waiting response from remote agency node.

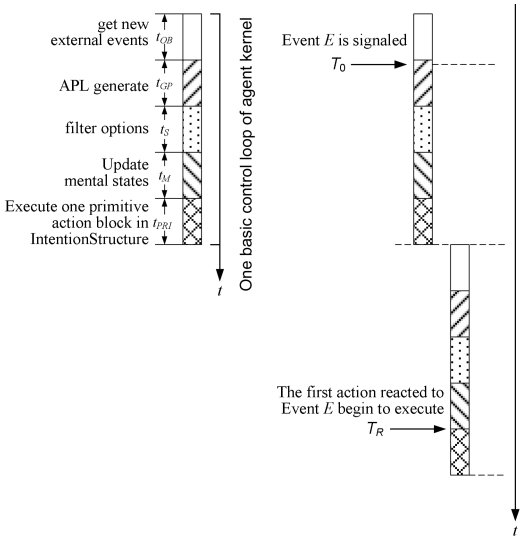


Fig. 7. Reactivity analysis of IA kernel

Table 1. Statistic values of agency node A2 running on the network management example

	Diagnosis-agent	Control-agent
Number of facts initiated and generated	mean: 70	mean: 84
Number of plans in the capacity library	12	16
Number of APLs generated	mean: 41	mean: 52
Average of APL generation time (second)	(mean: 0.016), (sd: 0.021)	(mean: 0.011), (sd: 0.013)
Average of NM-relevant plan execution time (second)	(mean: 0.2388), (max: 0.3154)	(mean: 0.4209), (max: 1.0263)

7 Related Work

Some attempts have been made to deal with the complexity in building adaptive and scalable management systems. For example, AI technologies (such as Expert system and Neural Network) [5, 8] are introduced to aid malfunction diagnose and add learning capacities. Many of classic intelligent agent systems [13, 15] belong to this category. Another trend is the use of distributed paradigm [14, 16], where the management goal is achieved via the cooperation of spatially distributed control authorities. Many multiagent systems [1, 4, 7] can be treat as this category. However, both of them have limitations: most systems of the former still follows centralized paradigm; the later concern little about how the autonomy of local authority is realized.

There are also some other cognitive agent structures based on folk psychology mental states. Some of them use cognitive concepts to model and execute by explicitly manipulating on those mental structures [6]. However, these kinds of system are criticized for the pool expression capability and execution efficiency so that are not practically used. dMARS [11] and JAM [9] use explicit representations of cognitive concepts, and the concepts are processed procedurally rather than via theorem proving, which is similar with the IA kernel of iSMAcy. However, both of them use ad hoc, not standard and open agent communication languages to interact, and concern little on how the multi-agency nodes are organized structurally.

8 Conclusion

Traditional AI methods focus on the internal reasoning and analysis process of single agent; whereas traditional distributed systems attend to the platform infrastructure that support multiagent communication and coordination. The iSMAcy system proposed here is trying to integrate these two methods into a coherent system. On one hand, it is a distributed management system with each agency node run as a local controller. Agents send and receive messages through standard FIPA agent communication language. On the other hand, the application agent, such as the diagnosis agent and control agent follow an intentional agent kernel with which the agent can decide whether to expand current intentions to execute so that working in a goal-directed manner; or change its focus completely by purchasing new goals and plans so that working in a reactive manner.

References

1. Bhutani K. R., Khan B.: Optimal distribution of a hierarchy of network management agents. Information Sciences. 149 (2003) 235-248.
2. Calisti M.: FIPA Standards for promoting interoperability of industrial agent systems. Agentcities Infodays. FIPA Presentation, Helsinki Finland (2003).
3. Charlton P., Cattoni R.: Evaluating the deployment of FIPA standards when developing application services. International Journal of Pattern Recognition and Artificial Intelligence. 15 (2001) 551-577.

4. Chen W. E., Hu C. L.: A mobile agent-based active network architecture for intelligent network control. Information Sciences. 141 (2002) 3-35.
5. Cheung C. F., Lee W. B., Wang W. M., et al.: Multi-perspective knowledge-based system for customer service management. Expert Systems with Applications. 24 (2003) 457-470.
6. Fischer M.: A survey of Concurrent Metatem—the language and its applications. In: Gabbay D. M. and Ohlbach H. J. (editors):. Temporal Logic—Proceedings of the 1st International Conference (LNAI Vol. 827). Springer-Verlag, Berlin (1994) 480-505.
7. Gavalas D., Greenwood D., et al.: Hierarchical network management: a scalable and dynamic mobile agent-based approach. Computer Networks. 38 (2002) 693-711.
8. Huang H. Y., Yen J. Y., et al.: Development of an intelligent energy management network for building automation. IEEE Transactions on Automation Science and Engineering. 1 (2004) 14-25.
9. Huber M. J.: JAM: A BDI-theoretic mobile agent architecture. Proceedings of the Third International Conference on Autonomous Agents (Agents'99). Seattle (1999) 236-243.
10. Huhns M. and Singh M.: Cognitive agents. IEEE Internet Computing. 2 (1998) 87-89.
11. d'Inverno M., Kinny D., Luck M. and Wooldridge M.: A Formal Specification of dMARS. In: Singh M. P., Rao A. S. and Wooldridge M. (editors): Intelligent Agents IV (LNAI Vol. 1365). Springer-Verlag, (1998) 155-176.
12. Jonker C., Treur J. and de Vries W.: External attribution of intentional notions to explain and predict agent behaviour. In: Müller J., Andre E., et al, (editors): Proceedings of the Fifth International Conference on Autonomous Agents. ACM Press, (2001) 41-42.
13. Manvi S., Venkataram P.: Applications of agent technology in communications: a review. Computer Communications. 27 (2004) 1493-1508.
14. Shah N. B., Mohapatra S. K.: Integrating and managing converged multi-service networks. BELL LABS TECH J. 10 (2005) 139-156.
15. Ryu K., Son Y. and Jung M.: Modeling and specifications of dynamic agents in fractal manufacturing systems. Computers in Industry. 52 (2003) 161-182.
16. Yu W., Wang J.: Scalable network resource management for large scale Virtual Private Networks. Simulation Modeling Practice and Theory. 12 (2004) 263-285.

Hybrid Intelligent Aircraft Landing Controller and Its Hardware Implementation

Jih-Gau Juang and Bo-Shian Lin

Department of Communications and Guidance Engineering
National Taiwan Ocean University, Keelung 20224, Taiwan, China
jgjuang@mail.ntou.edu.tw

Abstract. The purpose of this paper is to investigate the use of hybrid intelligent control to aircraft automatic landing system. Current flight control law is adopted in the intelligent controller design. Tracking performance and adaptive capability are demonstrated through software simulations. Two control schemes that use neural network controller and neural controller with particle swarm optimization are used to improve the performance of conventional automatic landing system. Control gains are selected by particle swarm optimization. Hardware implementation of this intelligent controller is performed by DSP with VisSim platform. The proposed intelligent controllers can successfully expand the controllable environment in severe wind disturbances.

1 Introduction

According to Boeing's report [1], 67% of the accidents by primary cause are due to human factors and 5% are attributed to weather factors. By phase of flight, 47% accidents are during final approach or landing. It is therefore desirable to develop an intelligent automatic landing system (ALS) that expands the operational envelope to include safer responses under a wider range of conditions. In this study, robustness of the proposed controller is obtained by choosing optimal control gain parameters that allow a wide range of disturbances to the controller. In [2], Kennedy and Eberhart presented an evolutionary computation algorithm the real-coded particle swarm optimization (PSO). PSO is one of the latest population-based optimization methods, which dose not use filtering operation (such as crossover and mutation). Members of the entire population are maintained through the search procedure. This method was developed through the simulation of a social system, and it has been found to be robust in solving continuous nonlinear optimization problems [3-5]. The method is also suitable for determination of the control parameters which give aircraft better adaptive capability in severe environments.

In the research of intelligent flight control, neuro-control is the most used technique. Juang [6-7] had presented a sequential learning technique that uses conventional neural network with back-propagation through time algorithm in landing control successfully. Aircraft landing control based on software simulations has also been done by many researchers [8-13]. These researches are all software simulations. In this paper, we present two control schemes for hardware simulation, a neural network controller and a neural network controller with PSO algorithm. A conventional feedforward neural network controller is designed using VisSim software and C language. The controller is

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 972 – 981, 2006.

trained by backpropagation algorithm. To improve the performance of the automatic landing system, PSO algorithm is implemented to the controller scheme, which can search more suitable control gains of the pitch autopilot. Furthermore, we download the code to the target DSP chip, and run it on the VisSim platform. Finally, the comparisons of simulation results by software and hardware are presented.

2 Aircraft Simulation Model

VisSim is a Windows-based program for the modeling and simulation of complex nonlinear dynamic systems [14]. VisSim combines an intuitive drag & drop block diagram interface with a powerful simulation engine. The visual block diagram interface offers a direct method for constructing, modifying and maintaining system models. The simulation engine provides fast and accurate solutions for linear, nonlinear, continuous time, discrete time, time varying and hybrid system designs. In here, we build a aircraft dynamic model under VisSim software, and realize the conventional PI controller by the same manner. Then, the intelligent controller design using C language and its realization by DSP are presented. Fig. 1 describes the aircraft landing system using conventional PI controller in VisSim platform.

Since the invention of the transistor and integrated circuit, digital signal processing functions have been implemented on many hardware platforms ranging from special-purpose architectures to general-purpose computers. It was not until all of the functionality (arithmetic, addressing, control, I/O, data storage, control storage) could be realized on a single chip that DSP could become an alternative to analog signal processing for the wide span of applications that we see today. In this study, we use TI TMS320LF2407 chip to perform our task. The 2407A devices offer the enhanced TMS320DSP architectural design of the C2xx core CPU for low-cost, low-power, and high-performance processing capabilities. Moreover, it offers suitable array of memory sizes and peripherals tailored to meet the specific performance points required by various applications. The TMS320LF2407 operates at 40 MHz (40 MIPS), has 4 to16 PWM output channels and has serial communication capabilities. In addition, the TMS320LF2407 contains a 10-bits analog-to-digital converter (ADC) having a minimum conversion time of 500 ns that offers up to 16 channels of analog input. Furthermore, the working process and externals of the eZdspTMLF2407A board are shown in Fig. 2 and Fig. 3, respectively.

There are three basics steps in the development of a DSP algorithm: 1). Create the system you want to execute on the target DSP; 2). Generate the C source code from the system; 3). Compile and link the C source code to produce an executable file. If step 1 performed using available blocks form VisSim software, thus, steps 2 and 3 are automatically performed by VisSim. But most intelligent methods do not exist in VisSim toolbox. So, an intelligent controller design using C language and its realization with DSP are presented in next section. Performance of a conventional PI controller that uses available blocks, as shown in Fig. 1, is presented in Table 1. Detail description of the aircraft model and control structure can be found in [7]. The controller can only guide the aircraft flying through wind speed of 0 ft/sec to 23 ft/sec. With the wind speed at 20 ft/sec, the horizontal position at touchdown is 891.2675 ft, horizontal velocity is 234.7 ft/sec, vertical speed is -2.7931 ft/sec, and pitch angle is -0.6848 degrees, as shown in Fig. 4 to Fig. 7.

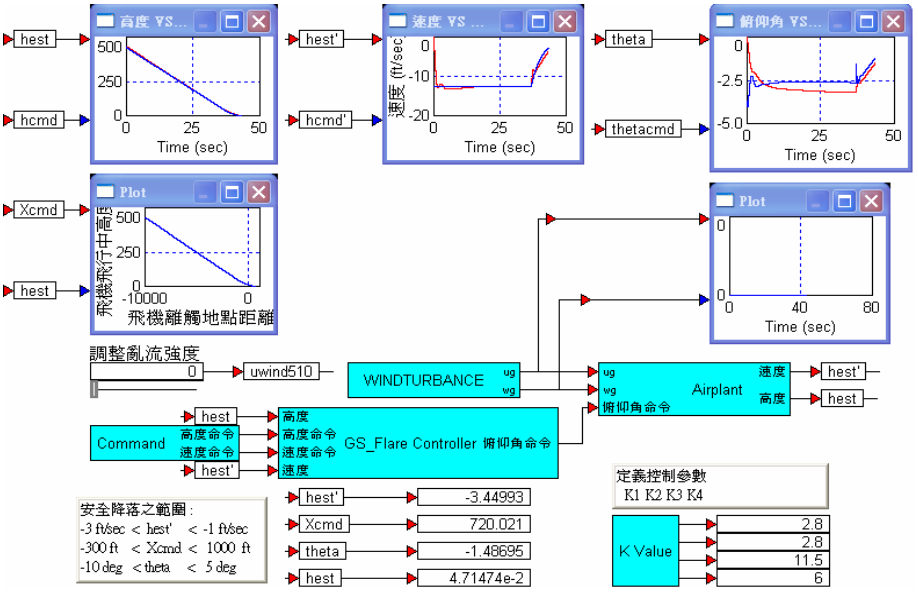


Fig. 1. Aircraft landing system using VisSim software

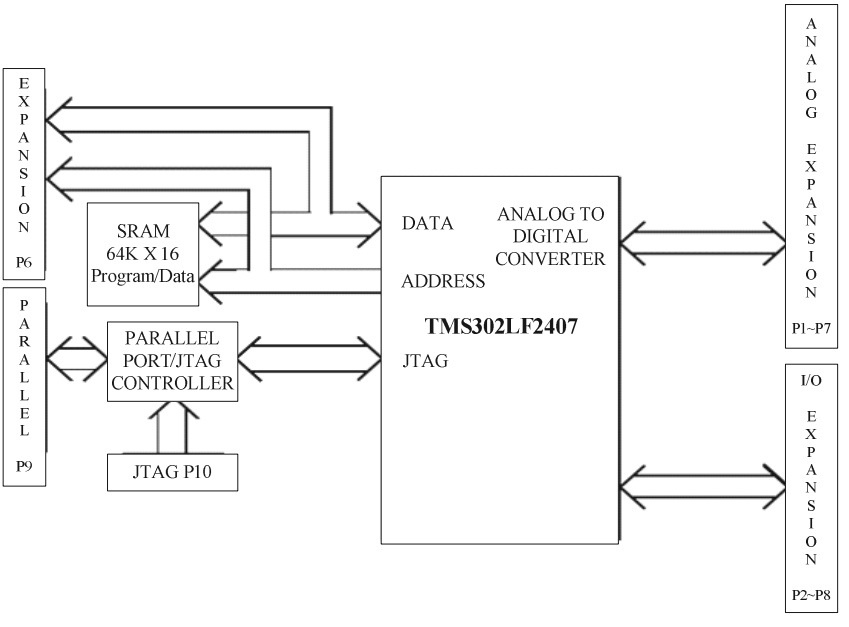


Fig. 2. Working process of the eZdspTMLF2407A board

Fig. 3. Externals of the eZdspTMLF2407A board

Table 1. The results from using conventional controller under VisSim & DSP platform (k1=2.8; k2=2.8; k3=11.5; k4=6.0)

Wind speed	Landing point (ft)	Aircraft vertical Speed (ft/sec)	Pitch angle (degree)
20	891.2675	-2.7931	-0.6848
23	977.0703	-2.4430	-0.4577
24	1000.7162	-2.4034	-0.4098

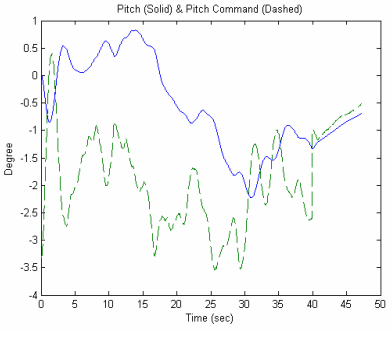


Fig. 4. Aircraft pitch and pitch command

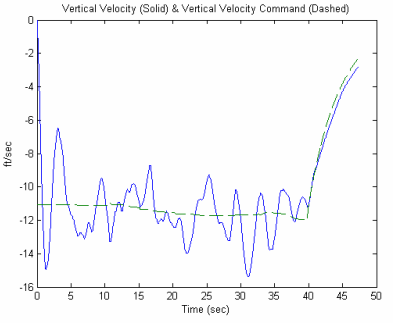


Fig. 5. Aircraft vertical velocity and velocity command

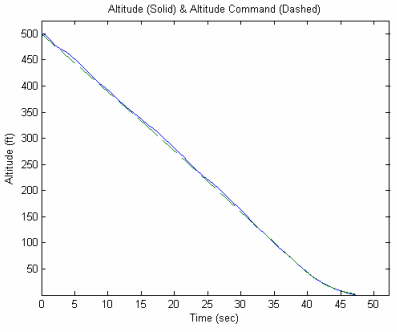


Fig. 6. Aircraft altitude and command

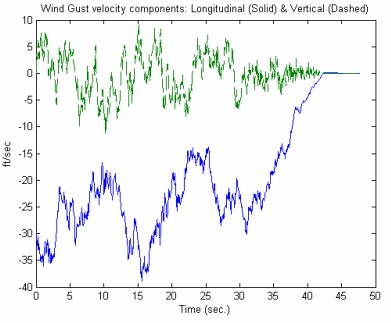


Fig. 7. Turbulence profile

3 Implementation of Intelligent Controller

At the beginning, the feedforward neural network controller is used and simulated on the Matlab simulation software. The controller is trained by backpropagation algorithm. The weight values of the controller are obtained after the BP training process (see Fig. 8). In here, the neural controller consists of four hidden units and three layers. The inputs for the neural controller are: altitude, altitude command, altitude rate, and altitude rate command. The output of the controller is the pitch command. Table 2 and Fig. 9 and Fig. 10 show the results from using different turbulence speeds. The controller can successfully guide the aircraft flying through wind speeds to 64 ft/sec.

After neural network training, parameters of control gains are selected by particle swarm optimization. With PSO, the controller can successfully guide the aircraft flying through wind speeds to 78 ft/sec. Table 3 shows the results from using different turbulence speeds. Fig. 11 and Fig. 12 show the results from using feedforward neural controller with the PSO.

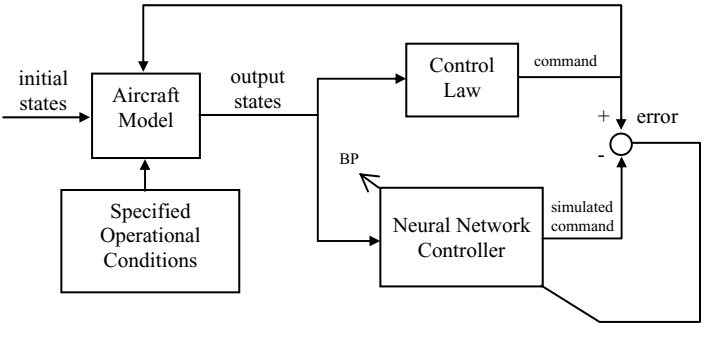


Fig. 8. Training process by BP algorithm

Table 2. The results from using conventional feedforward neural network controller (k1=2.8; k2=2.8; k3=11.5; k4=6.0)

Wind speed	Landing point (ft)	Aircraft vertical Speed (ft/sec)	Pitch angle (degree)
30	863	-2.08	-0.13
40	788	-2.52	0.11
50	761	-2.68	0.64
60	758	-2.57	1.23
64	573	-2.84	0.65

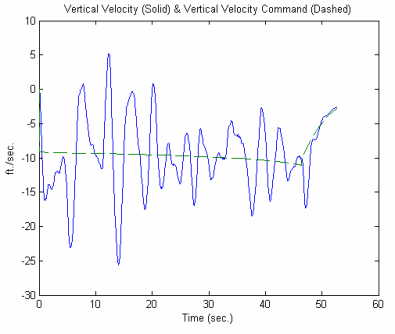


Fig. 9. Aircraft vertical velocity and velocity Command

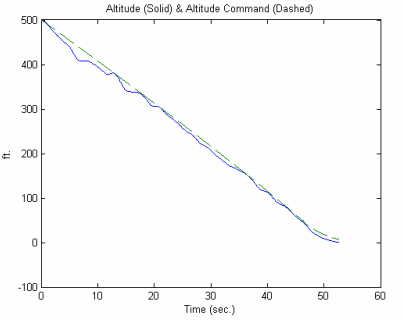


Fig. 10. Aircraft altitude and command

Table 3. The results from using feedforward neural network controller with PSO (k1=1.9351; k2=1.9351; k3=11.9771; k4=21.6547)

Wind speed	Landing point (ft)	Aircraft vertical Speed (ft/sec)	Pitch angle (degree)
30	880.3308	-2.3278	-0.3177
40	889.6822	-2.3897	-0.0410
50	759.6838	-2.3868	0.6300
60	918.6127	-2.3258	0.6614
70	844.9533	-2.6691	1.0621
78	252.2103	-2.9884	2.2176

To design a neural network controller block in VisSim, we utilize the C++ language and VisSim tools to implement the task. It is a complex procedure for making a special block. Fig. 13 shows a simple flowchart of the process.

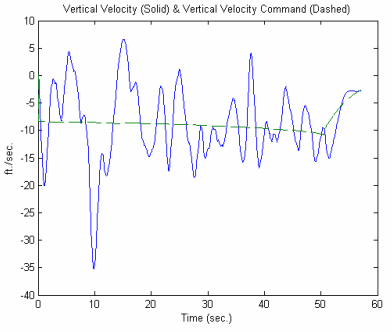


Fig. 11. Aircraft vertical velocity and velocity command

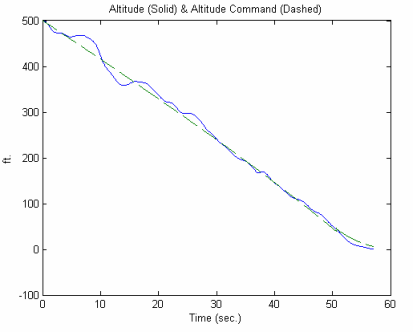


Fig. 12. Aircraft altitude and command

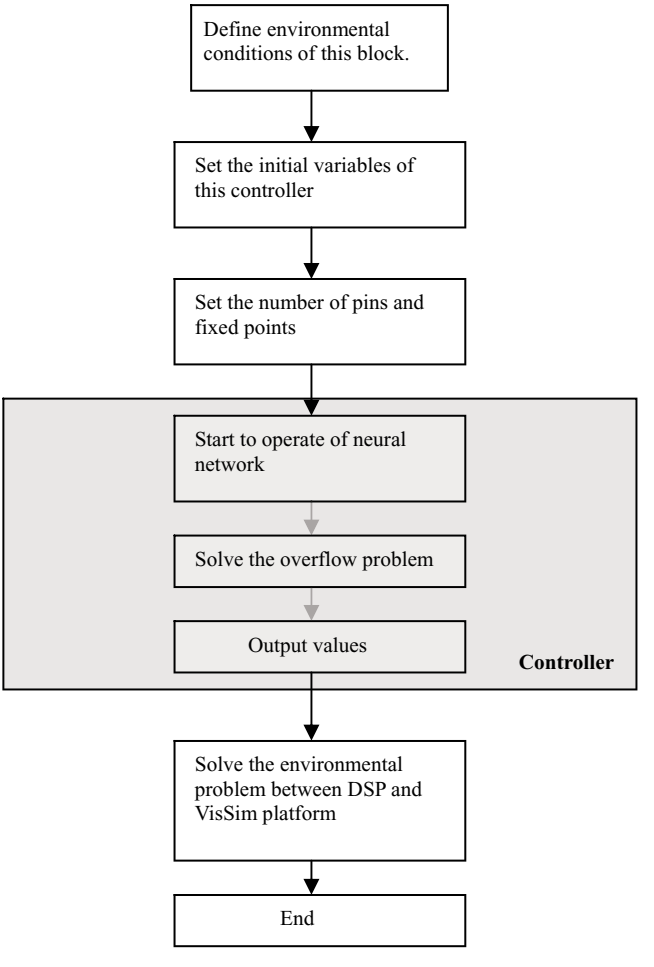


Fig. 13. Flowchart of designing process

Fig. 14 illustrates the aircraft landing system using feedforward neural controller on DSP chip and VisSim platform. DSP controller receives information form PC via printer port, and shows results on the VisSim platform. Table 4 shows the results from using different wind turbulence speeds. The controller can successfully guide the aircraft flying through wind speeds of 0 ft/sec to 53 ft/sec.

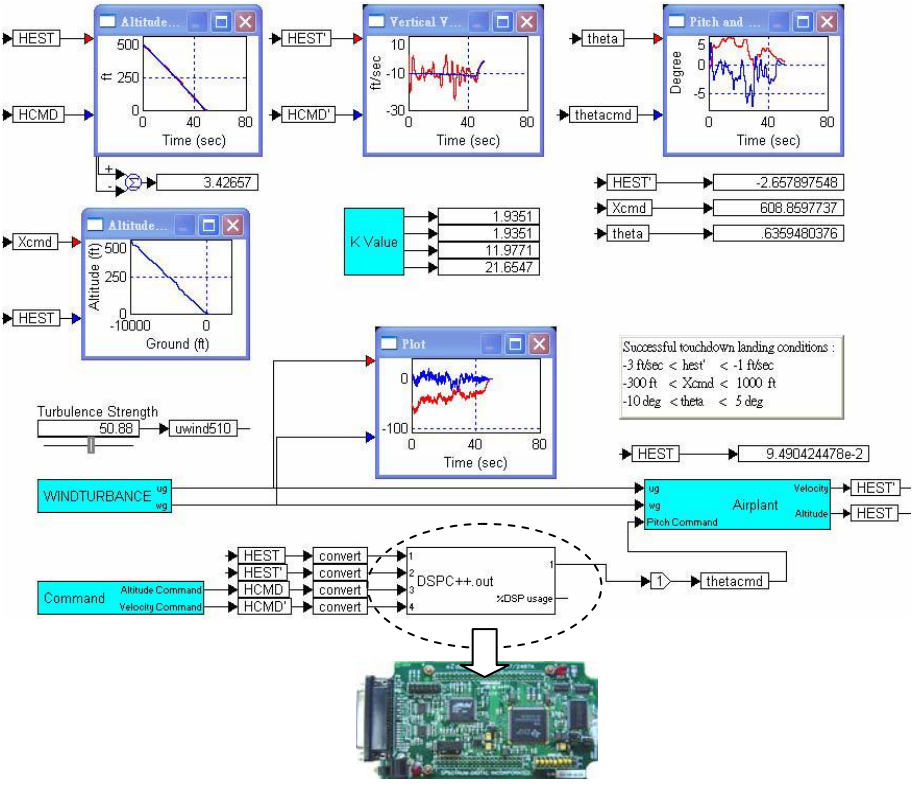


Fig. 14. DSP hardware-in-the-loop mode

Table 4. The results from using feedforward neural controller under VisSim & DSP (k1=2.8 ; k2=2.8; k3=11.5; k4=6.0)

Wind speed	Landing point (ft)	Aircraft vertical Speed (ft/sec)	Pitch angle (degree)
10	951.2347	-2.6283	-0.9725
20	866.0525	-2.7281	-0.6290
30	873.3354	-2.6465	-0.2101
40	857.5633	-2.4649	0.1952
50	787.6701	-2.2321	0.5977
53	905.3160	-2.0837	0.7025

Following presentation is to demonstrate the performance of the hardware controller with optimal control gains used in VisSim platform. These control gains are searched by the PSO, which can make the system more intelligent. Table 5 shows the results from using DSP with optimal control gains through VisSim. The controller can successfully overcome turbulence to 76 ft/sec.

Table 5. The results from using feedforward neural controller in DSP chip (k1=1.9351; k2=1.9351; k3=11.9771; k4=21.6547)

Wind speed	Landing point (ft)	Aircraft vertical Speed (ft/sec)	Pitch angle (degree)
10	881.7878	-2.7363	-1.0080
20	879.5342	-2.7743	-0.6337
30	845.1740	-2.7457	-0.2244
40	769.7975	-2.5618	0.2048
50	607.4361	-2.6658	0.6037
60	751.0171	-2.3735	1.0436
70	421.8852	-2.7395	1.7917
76	565.3009	-2.3718	1.8486

4 Conclusions

This paper presents an application of DSP chip for designing a feedforward neural network controller for aircraft landing system. Tracking performance and adaptive capability are demonstrated through software simulations. Compared with the software simulation results, hardware simulation of the DSP controller demonstrates the reliability of neural network method. On the other hand the PSO method improves the performance indeed. The PSO adopted in neural networks has the advantage of using fewer hidden neurons. This is because the PSO method can be used to generate high quality solutions on complex parameter searches. From simulations we know that the DSP controller is suitable for real implementation of online control. Although its performance is slightly worse than software simulation but it can overcome more disturbance range than previous studies. Simulation results show that the proposed automatic landing controllers can successfully expand the safety envelope of an aircraft to include severe wind disturbance environments without using the conventional gain scheduling technique.

Acknowledgement

This work was supported by the National Science Council, Taiwan under Grant NSC 92-2213-E-019 -005.

References

1. Boeing Publication: Statistical Summary of commercial Jet Airplane Accidents. Worldwide Operations 1959-1999 (2000)
2. Kennedy, J., Eberhart, R.C.: Swarm Intelligence. Morgan Kauffman publishers, San Francisco, CA (2001)
3. Kennedy, J., Eberhart, R.C.: Particle Swarm Optimization. Proceedings of IEEE International Conference on Neural Networks. 4 (1995) 1942-1948
4. Shi, Y., Eberhart, R.C.: Empirical Study of Particle Swarm Optimization. Proceedings of the 1999 Congress on Evolutionary Computation. (1999) 1945-1950
5. Angeline, P.J.: Using Selection to Improve Particle Swarm Optimization. Proceedings of IEEE International Conference on Evolutionary Computation. (1998) 84-89
6. Juang, J.G., Chang, H.H., Cheng, K.C.: Intelligent Landing Control Using Linearized Inverse Aircraft Model. Proceedings of American Control Conference. 4 (2002) 3269-3274
7. Juang, J.G., Chang, H.H., Chang, W.B.: Intelligent Automatic Landing System Using Time Delay Neural Network Controller. Applied Artificial Intelligence. 17 (2003) 563-581
8. Izadi, H., Pakmehr, M., Sadati, N.: Optimal Neuro-Controller in Longitudinal Autolanding of a Commercial Jet Transport. Proc. IEEE International Conference on Control Applications. CD-000202 (2003) 1-6
9. Chaturvedi, D.K., Chauhan, R., Kalra, P.K.: Application of generalized neural network for aircraft landing control system. Soft Computing. 6 (2002) 441-118
10. Iiguni, Y., Akiyoshi, H., Adachi, N.: An Intelligent Landing System Based on Human Skill Model. IEEE Transactions on Aerospace and Electronic Systems. 34 (1998) 877-882
11. Ionita , S., Sofron, E.: The Fuzzy Model for Aircraft Landing Control. Proc. AFSS International Conference on Fuzzy Systems. (2002) 47-54
12. Nho , K., Agarwal, R.K.: Automatic Landing System Design Using Fuzzy Logic. Journal of Guidance, Control, and Dynamics. 23 (2000) 298-304
13. Jorgensen , C.C., Schley, C.: A Neural Network Baseline Problem for Control of Aircraft Flare and Touchdown. Neural Networks for Control. (1991) 403-425
14. Visual Solutions, Inc.: VisSim User's Guide-Version 5.0. (2002)

Forecasting GDP in China and Efficient Input Interval

Cui Yu-quan[1], Ma Li-jie[1], and Xu Ya-peng[2]

[1] School of Mathematics and System Sciences, Shandong University, Jinan 250100, Shandong, China
cuiyq@sdu.edu.cn, cnqy129@126.com
[2] Shandong University of Science and Technology, Taian 271029, Shandong, China
xchh824@sina.com.cn

Abstract. In this paper we first give a new method of time series model to forecast GDP of China. The method proposed here aims to emphasis the importance of the impact of STEP-Affair on the GDP forecasting. The superiority of the method to ARMA model is illustrated in an example presented accordingly. Then in the system of whole economic when the GDP forecasted above is given, how can we allocate the limited resources to make the economical behavior relative efficient. We use data envelopment analysis to show how to determine input interval. Each input among the input interval as well as the given output constitute an efficient decision making unit (DMU). For decision makers the techniques are very important in decision making, especially in macroeconomic policies making in China.

1 Introduction

In macroeconomic a goal of long-term or short-term should be developed by policies makers to provide a sustainable development of economic. To attain the goal it is very important for a policy maker to predict the output of economic behavior and to allocate the limited resources. In this paper we present a technique to forecast GDP in China and then when GDP is given, we give a method of determining input interval. The inputs determined by our method and the given GDP constitute efficient DMUs so as to attain our goal.

The Gross Domestic product in China is affected by many factors such as economic system, policy, scientific level and nature, etc al. There are many researches on GDP forecasting as well as per-person GDP forecasting [1,2,3]. Before 1978, China was one of the poorest countries in the world; the GDP per worker in China then was just as 2% of the GDP per worker in USA. However, after 1978, the economy of China has enjoyed a rapid growth period, which has created a surprising economic miracle to the whole world. Preliminary estimates by China National Bureau of Statistics (NBS) show that the gross domestic product (GDP) of China in 2000 was 8.9404 trillion yuan, an increase of 8 percent over the previous year in constant terms. This estimated growth rate was 0.9 percentage points higher than that in 1999.In this paper we focus on the reform and opening in China to propose a time series model about GDP variable, then predict the GDP from 2001to 2004, and then compare them with

L. Jiao et al. (Eds.): ICNC 2006, Part II, LNCS 4222, pp. 982 – 990, 2006.

the real observations. At last we forecast the GDP of 2006 (the first yea of 'the eleventh 5 year plan'). Thus we can obtain more valuable information for the whole 'the eleventh 5' plan.

Data envelopment analysis (DEA) is a linear programming-based technique for evaluating the relative efficiency of a decision-making unit (DMU) by comparing it with other DMUs that utilize the same resources (inputs) to produce the same outputs. Since the first introduction of the technique by Charnes et al. [4], known as the CCR model, it has been extended and widely employed to assess performance of organization in various areas of public and private sectors, such as, schools, hospitals, banks districts, etc.[5]. After the GDP of short term or long term is predicted, how we can readjust the input of whole economics for the sake of the limited resources so as to make the input-output of the whole economical system to be efficient is a very important issue. It belongs to resource allocation in main macroeconomic. Here we give the interval of input estimation by using data envelopment analysis.

2 Conventional GDP Forecasting

Forecasting refers to a process by which the future behavior of a dynamical system is estimated based on your understanding and characterization of the system. More recently, soft computing methodologies, such as neural networks, fuzzy logic, and genetic algorithms, have been applies to the problem of forecasting complex time series. These methods have shown clear advantages over the traditional statistical ones. The advantage of soft computing methodologies is that we do not need to specify the structure of a model a priori, which is clearly needed in the classical regression analysis [6-9].

2.1 ARMA Model Review

Time series analysis has a wide use in handling the problem of forecasting[10-13]. Unfortunately, no simple statistical methodology like regression is known for dealing with covariance-nonstationarity. In addition, the presence of dependence in the data, whether of stationary or nonstationary kind, is problematic for someone who is used to analyzing observations. How does one deal with the dependent data? There are many ways enough to encourage a novice to venture into analyzing data and raising questions about the validity and prudence of the commonly used asymptotic procedures in finite sample situations. The idea of orthogonalization leads to powerful methodologies for handing stationary and nonstationarity data alike. Some of the most important techniques such as Box-Jenkins methods, the state-space and Kalman filtering algorithm, spectral analysis, the Wold decomposition and the Karhunen-Loeve expansion are centered and the notion of orthogonalizing segments of a stochastic process.

There are numerous ways to transform the dependent data x_t into a sequence of uncorrelated random variables called a white noise. Among these a central role is played by the Yule's(1927) and pioneering idea of auto-regression and its evolution into the autoregressive and moving average(ARMA) models of orders p and q:

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + \varepsilon_t + \sum_{j=i}^{q} \beta_j \varepsilon_{t-j}$$

Denoted by ARMA (p,q) for short. The goal of ARMA model fitting process is to use the data and develop an empirical method for choosing the orders p, q and estimating the parameters so that the fitted residuals are approximately uncorrelated. More precisely, the following three stages are used in the iterative ARMA model fitting process:

1 model identification (choosing p and q)
2 parameter estimation
3 diagnostic checking

As we know differencing and smoothing are two powerful and closely related methods for eliminating the trend in the data. However, at times it is preferred to estimate the nonconstant mean (trend) of the series while studying its dependence or covariance structure. This happens, for example, in economics and environmental sciences where estimation and testing for trend in presence of correlated data something takes the center stage. Verification of the global warming hypothesis and testing whether the national economy is in the growth (expansion) phase are some of the most prominent cases in these areas. Fortunately, this and the more problem of relating and explanatory (exogenous) variables to a time series of interest can be solved within the familiar linear regression framework.

2.2 A New Time Series Model with Step- Affair on the GDP in China

Unlike classical statistics, time series analysis is concerned with statistical inference from data that are not necessarily independent and identically distributed (i. i. d.). The ideal framework for time series analysis is stationarity whereas the data encountered in practice are often nonstationary for a variety of reasons such as trend, periodical factors, seasonal factors and random ones.Thus, the challenges facing a time series analyst are to transform the data to stationarity first, and then transform stationary data into i.i.d. sequence. We know a discrete-time stochastic process representing the evaluation of a system over time could be nonstationary because either its mean or its dependence as gauged by the covariance between two measurements is a function of time. For a given time series data, separating or distinguishing these two aspects of nonstationarity known as mean-nonstatinonarity and covariance-nonstationarity, respectively, is a difficult task. Traditionally, however, much attention is paid to correcting the mean nonstationarity by relying on simple transformations such as logarithm, differencing and smoothing. When confronted with nonstationarity, it is prudent to find simple transformations of the data to reduce the degree of nonstationarity. Since most time series of growth phenomena have monotone trend and variability, taking logarithm, successive differences and some other linear operations are the simplest and most useful transformations available to time series data analysts for eliminating trends, seasonality, variability and reducing data to stationarity.

In the following part a new time series analysis is proposed according to the situation of China in GDP forecasting. Table 1 shows the GDP in china from 1952 to 2000. Of course the data are nonstationarity and include upward trends; differencing is a powerful method for elimination the trend in the data (see figure 2).

Table 1. GDP in China from 1952 to 2000 (billion RMB)

Year	GDP	Y	GDP	Y	GDP	Y	GDP	Y	GDP
1952	67.9	1962	115	72	251.8	82	529.5	1992	2664
1953	82.9	1963	123	73	272.1	83	593.5	1993	3463
1954	85.9	1964	145	74	279.0	84	717.1	1994	4676
1955	197	1965	172	75	299.7	85	896.4	1995	5848
1956	103	1966	187	76	294.4	86	1020	1996	6789
1957	107	1967	177	77	320.2	87	1196	1997	7446
1958	131	1968	172	78	362.4	88	1493	1998	7835
1959	144	1969	194	79	403.8	89	1691	1999	8191
1960	146	1970	225	80	451.8	90	1855	2000	8940
1961	122	1971	243	81	486.2	91	2162		

(Data from China Statistical Yearbook 2004)

Note: for convenience, Year is denoted as Y and 72-91 denotes 1972-1991.

Table1 demonstrates that the GDP in China has increased from 362.4 (billion RMB) to 8940.4(billion RMB) in the period between 1978 and 2000, which is nearly thirty times in increase. From 1978, Chinese economy has been rocketing as a rapid speed. However, there is still large difference of the GDP compared with developed countries. The following graphs show the GDP and its one-order differencing GDP in China.

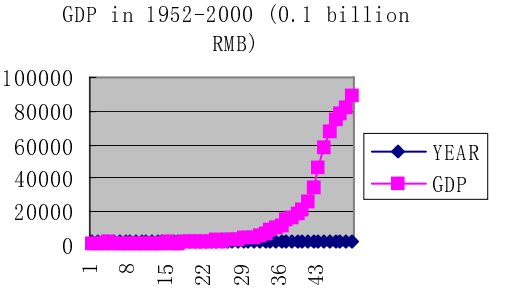


Fig. 1. GDP in 1952-2000 (0.1 billion RMB)

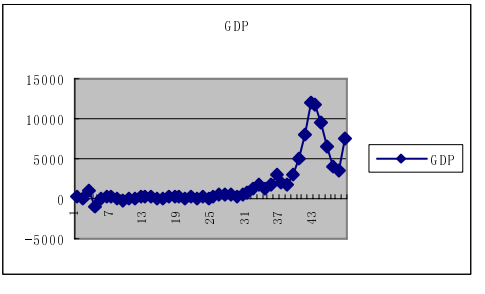


Fig. 2. GDP data of one-order differencing

Figure 1 shows the growth rate of GDP from 1952 to 2000. Since 1978 the average annual growth rate of GDP had reached about 9.4%, which was nearly as twice more as the average level in the whole world. Further, the figure demonstrates that there are four growth peaks within the 23 years, which are separately around the years of 1978, 1984, 1992, 2001. These four growth peaks might result from the four significant economic innovations in Chinese history: the rural economic system innovation began in 1978; the urban economic system innovation began in 1984; the policy of encouraging some areas developing first and then driving the dropped behind regions around 1992.

In the past twenty years, China has written a fantastic story of economic growth in the world. In figure 2, we can see that there is an upward trend in the acceleration of GDP since 1994. This is because GDP increased by 20% , in the following years the GDP even increased by 35% at most. In fact in time series analysis, some affairs happened suddenly, but its impact lasts. Among these affairs one is step-affair, another is pulse-affair, for the former, its impact will last for a long time after it takes place. So among the GDP from 1952 to 2000, the impact of economic reform and opening is a step-affair.

Let $x_t = \begin{cases} 0, t < u \\ 1, t \geq u \end{cases}$, where x_t is a dummy variable. Its impact will last for a long time after it takes place. So we formulate the model as follows:

$$\triangle y_t = a + \sum_{i=1}^{p} b_i \triangle y_{t-i} + cx_t + \varepsilon_t$$

In figure 2, there are altogether 48 points among which the reform and opening year of 1978 is the 26th point. So the GDP-forecasting model is:

$$\triangle y_t = a + \sum_{i=1}^{p} b_i \triangle y_{t-i} + cx_t + \varepsilon_t ,$$

Where $x_t = \begin{cases} 0, t < 26 \\ 1, t \geq 26 \end{cases}$, and $\triangle y_t = y_t - y_{t-1}$, ε_t is white noise. Having obtained observations of GDP, our goals of time series regression are: to estimate parameter a, b_i, c as well as the order p of the given model and to identify the nature of dependence of ε_t and to predict the future value y_{n+1}. Here we use FPE rule that is defined as follows: FPE= $\frac{N+p}{N-p}\sigma_n^2$. Where N is the number of $\{\triangle y_t\}$, p is the order of the model, and σ_n^2 is the residuals to the model. P is just the order of the model when its value makes FPE value the smallest. In details, we have 4 steps:

(1) Select P=1, 2, 3, …, using U-C estimation to estimate the parameter a, b_i, c

(2) Compute $\sigma_n^2 == E[a_t^2] = \frac{1}{N}\sum_{t=p+1}^{N} (x_t = b_1 \Delta y_{t-1} + b_2 \Delta y_{t-2} + ... + b_p \Delta y_{t-p})^2$

(3) Compute FPE= $\frac{N+p}{N-p}\sigma_n^2$.

(4) Compute different values of p, determining p and a, b_i, c to make the FPE smallest.

At last we obtain p=2, a =182.9863, b_1 =1.4723, b_2 =-0.6389, c =638.4273

So the model can be formulated:

$$\Delta y_t = 182.9863 + 1.4723\Delta y_{t-1} - 0.6389\Delta y_{t-2} + 638.4273x_t + a_t$$

We use the model stated above to forecast the GDP from 2000 to 2006 respectively. Comparing these with the real GDP, we obtain the relative errors (figure 2):

$y_{2001-2000}$ =9575 $y_{2002-2001}$ =12461, $y_{2003-2002}$ =11102, $y_{2004-2003}$ =11099,

$y_{2005-2004}$ =13097 , $y_{2006-2005}$ =11419

$\hat{y}_{2001} = y_{2000}$ +9575=98979 , $\hat{y}_{2002} = y_{2001}$ +12461=113022,

$\hat{y}_{2003} = y_{2002}$ +11102=123487 , $\hat{y}_{2004} = y_{2003}$ +11099=135596,

$\hat{y}_{2005} = y_{2004}$ +13097=151188, $\hat{y}_{2006} = \hat{y}_{2005}$ +11419=162609;

The relative errors: $\eta_{2001} = \frac{\hat{y}_{2001} - y_{2001}}{y_{2001}} = -1.57\%$, $\eta_{2002} = \frac{\hat{y}_{2002} - y_{2002}}{y_{2002}} = 0.57\%$,

$\eta_{2003} = \frac{\hat{y}_{2003} - y_{2003}}{y_{2003}} = -0.81\%$, $\eta_{2004} = \frac{\hat{y}_{2004} - y_{2004}}{y_{2004}} = -1.81\%$;

Table 2. Real GDP and GDP forecasting (10 billion RMB)

Year	Real GDP	Forecasting GDP	ARMA forecasting GDP	Relative errors η	ARMA relative errors
2001	100561	98979	95933	-1.57%	-4.6%
2002	112385	113022	102398	0.57%	-8.89%
2003	124497	123487	116694	-0.81%	-6.27%
2004	138091	135596	136515	-1.81%	-1.14%
2005		151188	153140		
2006		162609			

Table 2 shows the real GDP and the forecasting value of GDP from 2001 to 2006 by ARMR model and the model proposed in this paper. The results are satisfactory which illustrate the priority of our model to ARMA model.

By using the method proposed in this paper and ARMA model we got the GDP forecasting respectively. Figure 3 and figure 4 shows the comparison of GDP under the two forecasting methods and the real GDP from 2001 to 2005.

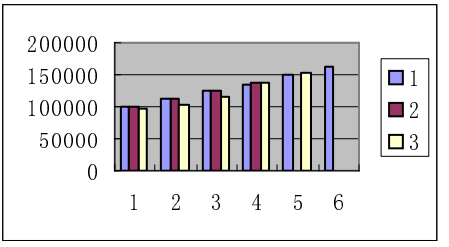


Fig. 3. 1 GDP forecasted by the method proposed here, 2 Real GDP, 3 GDP forecasting by using ARMA model

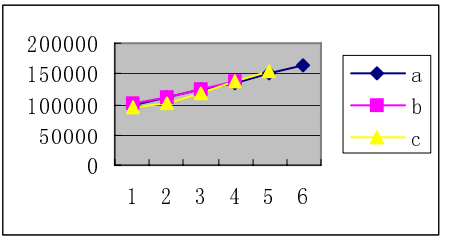


Fig. 4. a: GDP forecasted by the method proposed here, b: Real GDP, c: GDP forecasting by using ARMA model

Seen from figure 3 and figure 4, we know that the acceleration of GDP in China from 2001 to 2005 is 10.7%, 12.3%, 9.9%, 8.9%, 9.5% respectively which is in accord with the reports of annual economics of China government. According to this speed, the economics in China has step into a period with healthy stable development.

3 Determination of Efficient Input Interval

Data envelopment analysis is commonly used to determine the relative efficiency of peering decision-making units with multi-inputs and multi-outputs. Suppose the output vector of a DMU is given, how can we determine the interval of inputs so as to each input within the interval to make the DMU (X, Y) efficient?

Definition for n DMUs and a given output y_{n+1}, if $(x_{n+1,}, y_{n+1})$ is efficient (C^2GS^2) relative to n+1 DMUs and the efficient DMUs relative to n DMUS is still efficient. Then x_{n+1} is called efficient input of DMU_{n+1}.

Suppose y_{n+1} is a given number. For convenience, Let the proportion of the component of x_{n+1} is known, i.e. $x_{n+1} = lx_0$. Where x_0 is obtained from experience and

inputs are free disposal. Let J_1 denote the set of all the efficient DMUs relative to $DMU_j\ (j=1,2,\ldots n)$.

Let $h=\max\{h_{j'}\,|\,j'\in J_1\}$, where $h_{j'}$ is the optimal solutions of model $(\bar{G}_r)$ (seen in the following Theorem)

Theorem. For $DMU_{n+1}(l_0x_0, y_{n+1})$,

(1) If $l_0\geq h$, then for any $j'\in J_1$, $DMU_{j'}$ is efficient.

(2) If $l_0<h$,then there at least exists one $j'\in J_1$, $DMU_{j'}$ is inefficient.

Proof. For any $j'\in J_1$, then: $\begin{cases} u^{*T}x_{j'}-v^{*T}y_{j'}-\delta^*=0 \\ u^{*T}x_j-v^{*T}y_j-\delta^*=0,\ j\in J,\ j\neq j' \\ u^{*T}(l_0x_0)-v^{*T}y_{n+1}-\delta^*=0 \end{cases}$. Here

$u^*, v^*, \delta^*, h_{j'}$ is the optimal solution to $(\bar{G}_r)$: $\begin{cases} \min v^T y_{n+1} \\ u^{*T}x_{j'}-v^{*T}y_{j'}-\delta^*=0 \\ u^{*T}x_j-v^{*T}y_j-\delta^*\geq 0 \\ u^Tx_0=1 \\ u^T\geq \varepsilon e_1^{\ T}, v^T\geq \varepsilon e_2^{\ T} \end{cases}$. Since

$l_0\geq h\geq h_{j'}$, then $u^{*T}(l_0x_0)-v^{*T}y_{n+1}-\delta^*\geq 0$, that is to say, each decision-making unit in J_1 is still efficient for n+1 *DMUs*.

(1) Let $h=h_r\ r\in J_1$, when $l_0<h$, consider the efficiency score of DMU_r (C^2GS^2) relative to $n+1$ *DMUs*

Consider model (D_r) : $\begin{cases} \max v^T y_r+\pi \\ \omega^T x_j-\mu^T y_j-\alpha\geq 0,\ j=1,2,\ldots n \\ \omega^T(l_0x_0)-\mu^T y_{n+1}-\alpha\geq 0 \\ \omega^T x_r=1 \\ \omega^T\geq \varepsilon e_1^{\ T}, \mu^T\geq \varepsilon e_2^{\ T} \end{cases}$

If DMU_r is efficient for n+1 *DMUs*, then the optimal solution to model (D_r) denoted as $\omega^*, \mu^*, \alpha^*$ is certainly subject to $\mu^{*T}y_r+\alpha^*=1$ i.e. $\omega^{*T}x_r-\mu^{*T}y_r-\alpha^*=0$, for any $j\in J, j\neq r$, we have $\omega^{*T}-\mu^{*T}y_j-\alpha^*\geq 0$, So $\omega^*, \mu^*, \alpha^*$ must be the feasible solution to model $(\bar{G}_r)$ .Since $\omega^{*T}(l_0x_0)-\mu^{*T}y_{n+1}-\alpha^*\geq 0$, then $l_0\geq \dfrac{\mu^{*T}y_{n+1}+\alpha^*}{\omega^{*T}x_0}$.

However, $l_0<h=h_r=\min\dfrac{\mu^Ty_{n+1}+\alpha}{\omega^Tx_0}$. This is a contradiction, DMU_r is not

efficient. The lower limit of l_0 is obtained: $\underline{l} = h$. The upper limit of l_0 as well as g is obtained in the similar way.

So far we have obtained the lower limit and upper limit of l_0 on the condition that the output of one *DMU* and the proportion of inputs are given. Where $l_0 = [\underline{l}, \bar{l}] \in [h, g]$ and the corresponding input level is $[hx_0, gx_0]$.

4 Summary

In this paper we first proposed a new time series model in order to forecast the GDP in China. The prominent feature of the method is that it incorporated a factor called STEP-affair. This is because the economic reform and opening in 1978 has great impact on China' GDP. We use ARMA model and the technique proposed in the paper forecast the GDP in China respectively from 2001 to 2006. The results are very satisfactory which are in accordance with the situation in China. Then we give a method of determine an efficient input interval by using data envelopment analysis under the condition that when GDP is given.

References

1. ZHou Ming-lei.: Intervention Analysis with Time-Series to the World Crude Oil Price, Mathematics in practice and theory, Vol.134 (2004) 12-18
2. Wang Jian,: Appl ication of Temporal Sequence Analysis Technique in GDP Growth Prediction, Journal of Xiaogan University , Vol.25 (2005) 55-57
3. Hu Jin-hai, Xie Shou-sheng.: AR Model - based Prediction of Metal Content in Lubricating Oil, Experimentation and Research on Gas Turbine, Vol.16 (2003)32-35
4. A. Charnes, W.W. Cooper, E. Rhodes.: Measuring the efficiency of decision making units, European Journal of Operational Research, Vol.2 (1978) 429-444
5. A.Emrouznejad, E.Thanassoulis.: An extensive bibliography of data envelopment analysis, Volum Ⅰ and Ⅱ. Varwick Business School Research Paper, University of Warwick, 1996
6. Han, M., Xi, J., Xu, S., Yin, F.-L.: Prediction of chaotic time series based on the recurrent predictor neural network. IEEE Trans. Signal Processing, 52 (2004) 3409-3416
7. Wang, L.P., Teo, K.K., Lin, Z.: Predicting time series using wavelet packet neural networks. Proc. IJCNN 2001, 1593-1597
8. Castillo, O., Melin, P.: Hybrid intelligent systems for time series prediction using neural networks, fuzzy logic, and fractal theory. IEEE Trans. Neural Networks, 13 (2002) 1395-1408
9. Rajapakse, J.C., Wang, L.P. (Eds.): Neural Information Processing: Research and Development. Springer, Berlin (2004)
10. B.Abrahan, J.L.edolter.: Statistical Methods for Forecasting, Wiley, New York, 1983
11. T.W.Andereon.: The Statistical Analysis of Time Series, Wiley, New York, 1971
12. G.E.P.Box, G.M. Jekins , G.C.Reinsel.: Time Series Analysis- Forecasting and control, (Reised 3rd Edition), Prentice Hall, NJ1994
13. H.Helson, D.Lowdenslager.: Predition theory and Fourier series in several variables, Acta Mathematica 99 (1958)165-202

Author Index