

Jeng-Shyang Pan · Vaclav Snasel
Emilio S. Corchado · Ajith Abraham
Shyue-Liang Wang *Editors*

Intelligent Data Analysis and Its Applications, Volume 1

Proceeding of the First Euro-China
Conference on Intelligent
Data Analysis and Applications,
June 13–15, 2014, Shenzhen, China

Advances in Intelligent Systems and Computing

Volume 297

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

For further volumes:

<http://www.springer.com/series/11156>

About this Series

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within “Advances in Intelligent Systems and Computing” are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Advisory Board

Chairman

Nikhil R. Pal, Indian Statistical Institute, Kolkata, India
e-mail: nikhil@isical.ac.in

Members

Rafael Bello, Universidad Central “Marta Abreu” de Las Villas, Santa Clara, Cuba
e-mail: rbellop@uclv.edu.cu

Emilio S. Corchado, University of Salamanca, Salamanca, Spain
e-mail: escorchado@usal.es

Hani Hagrass, University of Essex, Colchester, UK
e-mail: hani@essex.ac.uk

László T. Kóczy, Széchenyi István University, Győr, Hungary
e-mail: koczy@sze.hu

Vladik Kreinovich, University of Texas at El Paso, El Paso, USA
e-mail: vladik@utep.edu

Chin-Teng Lin, National Chiao Tung University, Hsinchu, Taiwan
e-mail: ctlin@mail.nctu.edu.tw

Jie Lu, University of Technology, Sydney, Australia
e-mail: Jie.Lu@uts.edu.au

Patricia Melin, Tijuana Institute of Technology, Tijuana, Mexico
e-mail: epmelin@hafsamx.org

Nadia Nedjah, State University of Rio de Janeiro, Rio de Janeiro, Brazil
e-mail: nadia@eng.uerj.br

Ngoc Thanh Nguyen, Wroclaw University of Technology, Wroclaw, Poland
e-mail: Ngoc-Thanh.Nguyen@pwr.edu.pl

Jun Wang, The Chinese University of Hong Kong, Shatin, Hong Kong
e-mail: jwang@mae.cuhk.edu.hk

Jeng-Shyang Pan · Vaclav Snasel
Emilio S. Corchado · Ajith Abraham
Shyue-Liang Wang
Editors

Intelligent Data Analysis and Its Applications, Volume 1

Proceeding of the First Euro-China Conference
on Intelligent Data Analysis and Applications,
June 13–15, 2014, Shenzhen, China

Editors

Jeng-Shyang Pan
National Kaohsiung University of Applied
Sciences
Kaohsiung
Taiwan

Vaclav Snasel
Faculty of Elec. Eng. & Comp. Sci.
Department of Computer Science
VSB-Technical University of Ostrava
Ostrava-Poruba
Czech Republic

Emilio S. Corchado
Facultad de Biología
Departamento de Informática y Automática
University of Salamanca
Salamanca
Spain

Ajith Abraham
Scientific Network for Innovation and
Research Excellence
Machine Intelligence Research Labs
(MIR Labs)
Auburn Washington
USA

Shyue-Liang Wang
Department of Information Management
National University of Kaohsiung
Kaohsiung
Taiwan

ISSN 2194-5357

ISBN 978-3-319-07775-8

DOI 10.1007/978-3-319-07776-5

Springer Cham Heidelberg New York Dordrecht London

ISSN 2194-5365 (electronic)

ISBN 978-3-319-07776-5 (eBook)

Library of Congress Control Number: 2014940737

© Springer International Publishing Switzerland 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This volume composes the proceedings of the First Euro-China Conference on Intelligent Data Analysis and Applications (ECC 2014), which was hosted by Shenzhen Graduate School of Harbin Institute of Technology and was held in Shenzhen City on June 13–15, 2014. ECC 2014 was technically co-sponsored by Shenzhen Municipal People's Government, IEEE Signal Processing Society, Machine Intelligence Research Labs, VSB-Technical University of Ostrava (Czech Republic), National Kaohsiung University of Applied Sciences (Taiwan), and Secure E-commerce Transactions (Shenzhen) Engineering Laboratory of Shenzhen Institute of Standards and Technology. It aimed to bring together researchers, engineers, and policymakers to discuss the related techniques, to exchange research ideas, and to make friends.

113 papers were accepted for the final technical program. Four plenary talks were kindly offered by: Ljiljana Trajkovic (IEEE SMC president), C.L. Philip Chen (IEEE Fellow, University of Macau), Jhing-Fa Wang (Tajen University, Taiwan), and Ioannis Pitas (University of Thessaloniki, Greece).

We would like to thank the authors for their tremendous contributions. We would also express our sincere appreciation to the reviewers, Program Committee members and the Local Committee members for making this conference successful. Finally, we would like to express special thanks for the financial support from Shenzhen Municipal People's Government and Shenzhen Graduate School of Harbin Institute of Technology in making ECC 2014 possible.

June 2014

Jeng-Shyang Pan
Vaclav Snasel
Emilio S. Corchado
Ajith Abraham
Shyue-Liang Wang

Conference Organization

Honorary Chairs

Han-Chieh Chao National Ilan University, Taiwan

Advisory Committee Chairs

Tzung-Pei Hong National University of Kaohsiung, Taiwan
Bin-Yih Liao National Kaohsiung University of
Applied Sciences, Taiwan

Conference Chairs

Jeng-Shyang Pan Harbin Institute of Technology Shenzhen
Graduate School, China
Vaclav Snasel VSB-Technical University of Ostrava,
Czech Republic

Program Committee Chairs

Emilio S. Corchado University of Salamanca, Spain
Ajith Abraham Machine Intelligence Research Labs, USA
Shyue-Liang Wang National University of Kaohsiung, Taiwan

Invited Session Chairs

Wu-Chih Hu National Penghu University of Science and
Technology, Taiwan
Kuo-Kun Tseng Harbin Institute of Technology Shenzhen
Graduate School, China

Electronic Media Chairs

Jiun-Huei Ho	Cheng Shiu University, Taiwan
Tsu-Yang Wu	Harbin Institute of Technology Shenzhen Graduate School, China

Local Organizing Chairs

Yanfeng Zhang	Harbin Institute of Technology Shenzhen Graduate School, China
Chun-Wei Lin	Harbin Institute of Technology Shenzhen Graduate School, China
Chien-Ming Chen	Harbin Institute of Technology Shenzhen Graduate School, China

Publication Chairs

Shu-Chuan Chu	Flinders University, Australia
---------------	--------------------------------

Finance Chairs

Linlin Tang	Harbin Institute of Technology Shenzhen Graduate School, China
-------------	---

International Program Committee

Abdel hamid Bouchachia	University of Klagenfurt, Austria
Abd. Samad Hasan Basari	Universiti Teknikal Malaysia Melaka, Malaysia
Abraham Duarte	Universidad Rey Juan Carlos, Spain
Akira Asano	Kansai University, Japan
Alberto Alvarez	European Centre for Soft Computing, Spain
Alberto Cano	University of Cordoba, Spain
Alberto Fernandez	Universidad de Jaen, Spain
Alberto Bugarin	University of Santiago de Compostela, Spain
Alex James	Indian Institute of Information Technology and Management – Kerala, India
Alexandru Floares	Cancer Institute Cluj-Napoca, Romania
Alma Gomez	University of Vigo, Spain
Amelia Zafra Gomez	University of Cordoba, Spain
Amparo Fuster-Sabater	Institute of Applied Physics (C.S.I.C.), Spain
Ana Lorena	Federal University of ABC, Brazil
Anazida Zainal	Universiti Teknologi Malaysia, Malaysia
Andre Carvalho	University of Sao Paulo, Brazil

Andreas Koenig	Technische Universitat Kaiserslautern, Germany
Anna Bartkowiak	University of Wroclaw, Poland
Anna Fanelli	Universita di Bari, Italy
Antonio Peregrin	University of Huelva, Spain
Antonio J. Tallon-Ballesteros	University of Seville, Spain
Anusuriya Devaraju	Forschungszentrum Julich GmbH, Germany
Aranzazu Jurio	Universidad Publica de Navarra, Spain
Ashish Umre	University of Sussex, United Kingdom
Ashraf Saad	Armstrong Atlantic State University, United States
Ayeley Tchangani	University Toulouse III, France
Aymeric Histace	Universite Cergy-Pontoise, France
Azah Kamilah Muda	Universiti Teknikal Malaysia Melaka, Malaysia
Bartosz Krawczyk	Politechnika Wroclawska, Poland
Beatriz Pontes	University of Seville, Spain
Brijesh Verma	Central Queensland University, Australia
Carlos Barranco	Pablo de Olavide University, Spain
Carlos Cano	University of Granada, Spain
Carlos Fernandes	GeNeura Team, Spain
Carlos Garcia-Martinez	University of Cordoba, Spain
Carlos Lopezmolina	Universidad Publica de Navarra, Spain
Carlos Morell	Universidad Central Marta Abreu de Las Villas, Cuba
Cesar Hervás-Martínez	University of Cordoba, Spain
Chang-Shing Lee	National University of Tainan, Taiwan
Chao-Chun Chen	Southern Taiwan University, Taiwan
Chia-Feng Juang	National Chung-Hsing University, Taiwan
Chien-Ming Chen	Harbin Institute of Technology Shenzhen Graduate School, China
Chin-Chen Chang	Feng Chia University, Taiwan
Chris Cornelis	Ghent University, Belgium
Chuan-Kang Ting	National Chung Cheng University, Taiwan
Chu-Hsing Lin	Tunghai University, Taiwan
Chun-Wei Lin	Harbin Institute of Technology Shenzhen Graduate School, China
Coral del Val	University of Granada, Spain
Crina Grosan	Norwegian University of Science and Technology, Norway
Cristina Rubio-Escudero	University of Sevilla, Spain

Cristobal Romero	University of Cordoba, Spain
Cristobal J. Carmona	University of Jaen, Spain
Dalia Kriksciuniene	Vilnius University, Lithuania
David Becerra-Alonso	ETEA-INSA, Spain
Detlef Seese	Karlsruhe Institut of Technology (KIT), Germany
Edurne Barrenechea	Universidad Publica de Navarra, Spain
Eiji Uchino	Yamaguchi University, Japan
Eliska Ochodkova	VŠBTechnical University of Ostrava, Czech Republic
Elizabeth Goldbarg	Federal University of Rio Grande do Norte, Brazil
Emaliana Kasmuri	Universiti Teknikal Malaysia Melaka, Malaysia
Enrique Herrera-Viedma	University of Granada, Spain
Enrique Yeguas	University of Cordoba, Spain
Eulalia Szmidt	Systems Research Institute Polish Academy of Sciences, Poland
Eva Gibaja	University of Cordoba, Spain
Federico Divina	Pablo de Olavide University, Spain
Fernando Bobillo	University of Zaragoza, Spain
Fernando Delaprieta	University of Salamanca, Spain
Fernando Gomide	University of Campinas, Brazil
Fernando Jimenez	University of Murcia, Spain
Francesc J. Ferri	Universitat de Valencia, Spain
Francesco Marcelloni	University of Pisa, Italy
Francisco Fernandez Navarro	University of Cordoba, Spain
Francisco Herrera	University of Granada, Spain
Francisco Martinez-Alvarez	Pablo de Olavide University, Spain
Francisco Martinez-Estudillo	University Loyola Andalucia, Spain
Frank Klawonn	University of Applied Sciences Baunschweig, Germany
Gabriel Luque	University of Malaga, Spain
Gede Pramudya	Universiti Teknikal Malaysia Melaka, Malaysia
Giacomo Fiumara	University of Messina, Italy
Giovanna Castellano	Universita di Bari, Italy
Giovanni Acampora	University of Salerno, Italy
Girijesh Prasad	University of Ulster, United Kingdom
Gladys Castillo	University of Aveiro, Portugal
Gloria Bordogna	CNR IDPA, Italy
Gregg Vesonder	AT&T Labs Research, United States
Huiyu Zhou	Queen's University Belfast, United Kingdom
Ilkka Havukkala	Intellectual Property Office of New Zealand, New Zealand

Imre Lendak	University of Novi Sad, Serbia
Intan Ermahani A. Jalil	Universiti Teknikal Malaysia Melaka, Malaysia
Isabel Nunes	UNL/FCT, Portugal
Isabel S. Jesus	Instituto Superior de Engenharia do Porto, Portugal
Ivan Garcia-Magarino	Universidad a Distancia de Madrid, Spain
Jae Oh	Syracuse University, United States
Jan Martinovic	VŠB Technical University of Ostrava, Czech Republic
Jan Plato	VŠB Technical University of Ostrava, Czech Republic
Javier Perez	University of Salamanca, Spain
Javier Sedano	Technological Institute of Castilla y Leon, Spain
Jesus Alcala-Fdez	University of Granada, Spain
Jesus Serrano-Guerrero	University of Castilla-La Mancha, Spain
Jitender S. Deogun	University of Nebraska, United States
Joaquin Lopez Fernandez	University of Vigo, Spain
Jorge Nunez Mc Leod	Institute of C.E.D.I.A.C., Argentina
Jose Luis Perez de la Cruz	University of Malaga, Spain
Jose M. Merigo	University of Barcelona, Spain
Jose-Maria Luna	University of Cordoba, Spain
Jose Pena	Universidad Politecnica de Madrid, Spain
Jose Raul Romero	University of Cordoba, Spain
Jose Tenreiro Machado	Instituto Superior de Engenharia do Porto, Portugal
Jose Valente De Oliveira	Universidade do Algarve, Portugal
Jose Villar	Oviedo University, Spain
Juan Botia	Universidad de Murcia, Spain
Juan Gomez-Romero	Universidad Carlos III de Madrid, Spain
Juan Vidal	Universidade de Santiago de Compostela, Spain
Juan J. Flores	Universidad Michoacana de San Nicolas de Hidalgo, Mexico
Juan-Luis Olmo	University of Cordoba, Spain
Julio Cesar Nievola	Pontificia Universidade Catolica do Parana, Brazil
Jun Zhang	Waseda University, Japan
Jyh-Horng Chou	National Kaohsiung First Univ. of Science and Technology, Taiwan
Kang Tai	Nanyang Technological University, Singapore
Kaori Yoshida	Kyushu Institute of Technology, Japan
Kazumi Nakamatsu	University of Hyogo, Japan
Kelvin Lau	University of York, United Kingdom
Kubilay Ecerkale	Turkish Air Force Academy, Turkey

Kumudha Raimond	Karunya University, India
Kun Ma	University of Jinan, China
Leandro Coelho	Pontificia Universidade Catolica do Parana, Brazil
Lee Chang-Yong	Kongju National University, Korea
Leida Li	University of Mining and Technology, China
Leon Wang	National University of Kaohsiung, Taiwan
Liang Zhao	University of Sao Paulo, Brazil
Liliana Ironi	IMATI-CNR, Italy
Luciano Stefanini	University of Urbino “Carlo Bo”, Italy
Ludwig Simone	North Dakota State University, United States
Luigi Troiano	University of Sannio, Italy
Luka Eciolaza	European Centre for Soft Computing, Spain
Macarena Espinilla Estevez	Universidad de Jaen, Spain
Manuel Grana	University of Basque Country, Spain
Manuel Lama	Universidade de Santiago de Compostela, Spain
Manuel Mucientes	University of Santiago de Compostela, Spain
Marco Cococcioni	University of Pisa, Italy
Maria Nicoletti	Federal University of Sao Carlos, Brazil
Maria Torsello	Universita di Bari, Italy
Maria Jose Del Jesus	Universidad de Jaen, Spain
Mariantonietta Noemi La Polla	IIT-CNR, Italy
Maria Teresa Lamata	University of Granada, Spain
Mario Giovanni C.A. Cimino	University of Pisa, Italy
Mario Koeppen	Kyushu Institute of Technology, Japan
Martine De Cock	Ghent University, Belgium
Michael Blumenstein	Griffith University, Australia
Michal Kratky	VŠB Technical University of Ostrava, Czech Republic
Michal Wozniak	Wroclaw University of Technology, Poland
Michela Antonelli	University of Pisa, Italy
Mikel Galar	Universidad Publica de Navarra, Spain
Milos Kudelka	VŠB Technical University of Ostrava, Czech Republic
Min Wu	Oracle, United States
Noor Azilah Muda	Universiti Teknikal Malaysia Melaka, Malaysia
Norberto Diaz-Diaz	Pablo de Olavide University, Spain
Norton Gonzalez	University of Fortaleza, Brazil
Nurulakmar Emran	Universiti Teknikal Malaysia Melaka, Malaysia
Olgierd Unold	Wroclaw University of Technology, Poland
Oscar Castillo	Tijuana Institute of Technology, Mexico
Ovidio Salvetti	ISTI-CNR, Italy
Ozgur Koray Sahingoz	Turkish Air Force Academy, Turkey
Pablo Villacorta	University of Granada, Spain

Patrick Siarry	Universit de Paris, France
Paulo Carrasco	Universidade do Algarve, Portugal
Paulo Moura Oliveira	University of Tras-os-Montes and Alto Douro, Portugal
Pedro Gonzalez	University of Jaen, Spain
Philip Samuel	Cochin University of Science and Technology, India
Pierre-Francois Marteau	Universite de Bretagne Sud, France
Pietro Ducange	University of Pisa, Italy
Punam Bedi	University of Delhi, India
Qieshi Zhang	Waseda University, Japan
Qinghan Xiao	Defence R&D Canada, Canada
Radu-Codrut David	Politehnica University of Timisoara, Romania
Rafael Bello	Universidad Central de Las Villas, Cuba
Ramin Halavati	Sharif University of Technology, Iran
Ramiro Barbosa	Instituto Superior de Engenharia do Porto, Portugal
Ramon Sagarna	University of Birmingham, United Kingdom
Richard Jensen	Aberystwyth University, United Kingdom
Robert Berwick	Massachusetts Institute of Technology, United States
Roberto Armenise	Poste Italiane, Italy
Robiah Yusof	Universiti Teknikal Malaysia Melaka, Malaysia
Roman Neruda	Institute of Computer Science, Czech Republic
S. Ramakrishnan	Dr. Mahalingam College of Engineering and Technology, India
Sabrina Ahmad	Universiti Teknikal Malaysia Melaka, Malaysia
Sadaaki Miyamoto	University of Tsukuba, Japan
Santi Llobet	Universitat Oberta de Catalunya, Spain
Satrya Fajri Pratama	Universiti Teknikal Malaysia Melaka, Malaysia
Saurav Karmakar	Georgia State University, United States
Sazalinsyah Razali	Universiti Teknikal Malaysia Melaka, Malaysia
Sebastian Ventura	University of Cordoba, Spain
Selva Rivera	Institute of C.E.D.I.A.C., Argentina
Shang-Ming Zhou	University of Wales Swansea, United Kingdom
Shyue-Liang Wang	National University of Kaohsiung, Taiwan
Siby Abraham	University of Mumbai, India
Silvia Poles	EnginSoft, Italy
Silvio Bortoleto	Federal University of Rio de Janeiro, Brazil
Siti Rahayu Selamat	Universiti Teknikal Malaysia Melaka, Malaysia
Steven Guan	Xi'an Jiaotong-Liverpool University, China
Sung-Bae Cho	Yonsei University, Korea
Swati V. Chande	International School of Informatics and Management, India

Sylvain Piechowiak	Universite de Valenciennes et du Hainaut-Cambresis, France
Takashi Hasuike	Osaka University, Japan
Teresa Ludermir	Federal University of Pernambuco, Brazil
Thomas Hanne	University of Applied Sciences Northwestern Switzerland, Switzerland
Tsu-Yang Wu	Harbin Institute of Technology Shenzhen Graduate School, China
Tzung-Pei Hong	National University of Kaohsiung, Taiwan
Vaclav Snasel	VŠB Technical University of Ostrava, Czech Republic
Valentina Colla	Scuola Superiore Sant'Anna, Italy
Victor Hugo Menendez Dominguez	Universidad Autonoma de Yucatan, Mexico
Vincenzo Loia	University of Salerno, Italy
Vincenzo Piuri	University of Milan, Italy
Virgilijus Sakalauskas	Vilnius University, Lithuania
Vivek Deshpande	MIT College of Engineering, India
Vladimir Filipovic	University of Belgrade, Serbia
Wei Wei	Xi'an University of Technology, China
Wei-Chiang Hong	Oriental Institute of Technology, Taiwan
Wen-Yang Lin	National University of Kaohsiung, Taiwan
Wilfried Elmenreich	University of Klagenfurt, Austria
Yasuo Kudo	Muroran Institute of Technology, Japan
Ying-Ping Chen	National Chiao Tung University, Taiwan
Yun-Huoy Choo	Universiti Teknikal Malaysia Melaka, Malaysia
Yunyi Yan	Xidian University, China
Yusuke Nojima	Osaka Prefecture University, Japan

Contents

Part I: Data Security and Its Applications

A Robust Audio Zero-Watermarking Algorithm Based on Wavelet Packet Analysis	3
<i>Xueying Zhang, Wei Zhang, Fenglian Li, Guangyu Liu</i>	
Information Security Management for Higher Education Institutions	11
<i>Simon K.S. Cheung</i>	
Laser Induced Breakdown Spectroscopy Data Processing Method Based on Wavelet Analysis	21
<i>Lu Muchao</i>	
Towards Time-Bound Hierarchical Key Management in Cloud Computing	31
<i>Tsu-Yang Wu, Chengxiang Zhou, Eric Ke Wang, Jeng-Shyang Pan, Chien-Ming Chen</i>	
Shape Estimation from 3D Point Clouds	39
<i>Jingyong Su, Lin-Lin Tang</i>	
Deterministic Data Sampling Based on Neighborhood Analysis	47
<i>Sarka Zehnalova, Milos Kudelka, Jan Platos</i>	
Diagonal Interacting Multiple Model H_∞ Filtering for Simultaneous Sensor Localization and Target Tracking with NLOS Mitigation	57
<i>Xiaoyan Fu, Yuanyuan Shang, Hui Ding, Xiuzhuang Zhou</i>	

Part II: Intelligent Data Analysis and Its Applications

A Projection-Based Approach for Mining Highly Coherent Association Rules	69
<i>Chun-Hao Chen, Guo-Cheng Lan, Tzung-Pei Hong, Shyue-Liang Wang, Yui-Kai Lin</i>	

ICISLM: Design of an Integrated Cloud Information System for Logistic Management Based on Web Server Virtualization	79
<i>Shang-Liang Chen, Yun-Yao Chen, Hsuan-Pei Wang, Chiang Hsu</i>	
Hiding Sensitive Itemsets with Minimal Side Effects in Privacy Preserving Data Mining	87
<i>Chun-Wei Lin, Tzung-Pei Hong, Hung-Chuan Hsu</i>	
The Bridge Edge Label Propagation for Overlapping Community Detection in Social Networks	97
<i>Jui-Le Chen, Jen-Wei Hu, Chu-Sing Yang</i>	
A New Estimation of Distribution Algorithm to Solve the Multiple Traveling Salesmen Problem with the Minimization of Total Distance	103
<i>S.H. Chen, Y.H. Chen</i>	
Subspace Learning with Enriched Databases Using Symmetry	113
<i>Konstantinos Papachristou, Anastasios Tefas, Ioannis Pitas</i>	
Image Categorization Using Macro and Micro Sense Visual Vocabulary	123
<i>Chang-Ming Kuo, Chi-Kao Chang, Nai-Chung Yang, Chung-Ming Kuo, Yu-Ming Chen</i>	
Part III: Technologies for Next-Generation Network Environments	
An Incremental Algorithm for Maintaining the Built FUSP Trees Based on the Pre-large Concepts	135
<i>Chun-Wei Lin, Wensheng Gan, Tzung-Pei Hong, Raylin Tso</i>	
Another Improvement of RAPP: An Ultra-lightweight Authentication Protocol for RFID	145
<i>Xinying Zheng, Chien-Ming Chen, Tsu-Yang Wu, Eric Ke Wang, Tsui-Ping Chung</i>	
A Security System Based on Door Movement Detecting	155
<i>Ci-Rong Li, Chie-Yang Kuan, Bing-Zhe He, Wu-En Wu, Chi-Yao Weng, Hung-Min Sun</i>	
Network Performance QoS Prediction	165
<i>Jaroslav Frnda, Miroslav Voznak, Lukas Sevcik</i>	
Study on Security Analysis of RFID	175
<i>Yi Hou, Jialin Ma</i>	
Web Services Discovery with Semantic Based on P2P	181
<i>Jin Li, Yongyi Zhao, Bo Song</i>	

Analysis and Enhancement of TCP Performance in Ad Hoc Wireless Networks	189
<i>Li Miaoyan, Zhou Chuansheng</i>	

Part IV: Intelligent System Analysis and Social Networks

SGR-StarCraft: Somatosensory Game Rehabilitation via StarCraft	201
<i>Ching-Hsun Hsieh, Chia-Hui Wang</i>	

Discovering Sentiment of Social Messages by Mining Message Correlations	213
<i>Hsin-Chang Yang, Chung-Hong Lee, Chun-Yen Wu, Yu-Chian Huang</i>	

Seek the Consent, Respect the Dissent: An Analysis of User Behaviors in Online Collaborative Community	223
<i>Xiaoyue Tang, Hui Wang, Zhengzheng Ouyang, Wei Yu</i>	

Study on Parallax Scrolling Web Page Conversion Module	235
<i>Song-Nian Wang, Fong-Ming Shyu</i>	

A Graph Theory-Based Evaluation of Strategy Set in Robot Soccer	245
<i>Jie Wu, Václav Snášel, Guangzhao Cui</i>	

Spatial and Frequency Domain-Based Feature Fusion Method for Texture Retrieval	257
<i>Rurui Zhou</i>	

Comparisons of Typical Discrete Logistic Map and Henon Map	267
<i>Bingbing Song, Qun Ding</i>	

Part V: Intelligent Analysis for Biological, Mobile and Cloud Computing

Wavelet-Domain Image Watermarking Using Optimization-Based Mean Quantization	279
<i>Huang-Nan Huang, Der-Fa Chen, Chiu-Chun Lin, Shuo-Tsung Chen</i>	

The Sybil Attack in Participatory Sensing: Detection and Analysis	287
<i>Shih-Hao Chang, Kuo-Kun Tseng, Shin-Ming Cheng</i>	

Vessel Freeboard Calculation Method Based on Laser Scanning	299
<i>Yingce Zhao, Guangming Lu, Xiaotang Guo, Yazhuo Wang</i>	

Visual Information Analysis for Big-Data Using Multi-core Technologies	309
<i>Nikolaos Mpountouropoulos, Anastasios Tefas, Nikos Nikolaidis, Ioannis Pitas</i>	

Application of Job Shop Based on Immune Genetic Algorithm	317
<i>Lei Meng, Chuansheng Zhou</i>	

Study of Evaluation of GPS/BeiDou Combination Regional Navigation Satellite System 323
Tenghong Liu, Songlin Liu

Texture Image Classification Using Gabor and LBP Feature 329
Youfu Du

Part VI: Multimedia Innovative Computing

Effective Moving Object Detection from Videos Captured by a Moving Camera 343
Wu-Chih Hu, Chao-Ho Chen, Chih-Min Chen, Tsong-Yi Chen

Roadside Unit Deployment Based on Traffic Information in VANETs 355
Ji-Han Jiang, Shih-Chieh Shie, Jr-Yung Tsai

Overlapping Community Detection with a Maximal Clique Enumeration Method in MapReduce 367
Yi-Jen Su, Wei-Lin Hsu, Jian-Cheng Wun

Grey Analysis on Underwater Sensor Network of Penghu Set Net 377
Yih-Fuh Wang, Chang-Ling Tsai

A Research of Wireless Energy Collector for Increasing the Power of Rechargeable Device 383
Chuen-Ching Wang, Chi-Hung Wei

How to Determine the Best Indexes of Industry Website by FANP Approach 391
Chih-Chao Chung, Hsiu-Chu Huang, Huei-Yin Tsai, Shi-Jer Lou

Mobile Learning Achievement from the Perspective of Self-efficacy: A Case Study of Basic Computer Concepts Course 403
Yuh-Ming Cheng, Sheng-Huang Kuo, E-Liang Cheng

Part VII: Intelligent Technologies and Telematics Applications

The Implementation of OBD-II Vehicle Diagnosis System Integrated with Cloud Computation Technology 413
Jheng-Syu Zhou, Shi-Huang Chen

Daily Power Demand Forecast Models of the Differential Polynomial Neural Network 421
Ladislav Zjavka

Design of Embedded Ethernet Interface Based on ARM11 and Implementation of Data Encryption 431
Chunlei Fan, Zhiqiang Li, Qun Ding, Songyan Liu

The Evaluation of the Business Operation Performance by Applying Grey Relational Analysis	441
<i>Dingtao Zhao, Su-Hui Kuo, Tien-Chin Wang</i>	
An Echo-Aided Bat Algorithm to Construct Topology of Spanning Tree in Wireless Sensor Networks	451
<i>Yi-Ting Chen, Ming-Te Tsai, Bin-Yih Liao, Jeng-Shyang Pan, Mong-Fong Horng</i>	
Part VIII: Cross-Discipline Techniques in Signal Processing and Networking	
Design of Triple-Band Planar Dipole Antenna	465
<i>Yuh-Yih Lu, Jun-Yi Guo, Kai-Lun Chung, Hsiang-Cheh Huang</i>	
Solar Irradiance Estimation Using the Echo State Network and the Flexible Neural Tree	475
<i>Sebastián Basterrech, Tomáš Buriánek</i>	
A DOA Estimation Method for Wideband Signals with an Arbitrary Plane Array	485
<i>Jiaqi Zhen, Qun Ding, Bing Zhao</i>	
An e-Learning System Based on EGL and Web 2.0	495
<i>Xiaomei Li, Zhaozhe Ma, Bo Song</i>	
Adaptive Pulse Design and Spectrum Handoff Technology Based on Cognition	505
<i>Bing Zhao, Erfu Wang, Jiaqi Zhen, Qun Ding</i>	
Technology Research of the Configured Component ERP System Based on XML	515
<i>Jialin Ma, Yi Hou</i>	
Discriminative Feature Learning for Action Recognition Using a Stacked Denoising Autoencoder	521
<i>Ruoxin Sang, Peiquan Jin, Shouhong Wan</i>	
Author Index	533

Part I
Data Security and Its Applications

A Robust Audio Zero-Watermarking Algorithm Based on Wavelet Packet Analysis

Xueying Zhang, Wei Zhang, Fenglian Li, and Guangyu Liu

College of Information Engineering, Taiyuan University of Technology
Taiyuan, China
tyzhangxy@163.com

Abstract. This paper proposed a new robust audio zero-watermarking algorithm which can be used to authenticate the copyright of digital audio. This algorithm has the following features: (1) it extracts the low frequency components of original audio to construct zero-watermarking by using the wavelet packet analysis method, ensures the imperceptibility of watermarking algorithm; (2) it uses cubic spline interpolation and multilevel scrambling technology to construct a meaningful zero-watermarking with a binary text image; it improves its safety and the robustness toward the attacks. Meanwhile, it is fairly straightforward to finish the authentication. The experimental result shows that this algorithm has strong robustness against typical common attacks and hostile attacks.

Keywords: zero-watermark, robust watermarking, copyright authentication, cubic spline interpolation, wavelet packet analysis.

1 Introduction

As a popular research, digital audio watermarking technology[1],[2]has been widely used in many fields, such as copyright authentication[3], content authenticity, secure communication. At present, it is mainly depending on robust watermarking to realize the digital audio copyright authentication. According to the difference of forming mechanism, the robust watermarking can be divided into two types: embedded watermarking[4] and zero-watermarking[5]. Zero-watermarking is a typical digital watermarking system. It solved the conflict between the imperceptibility and robustness of digital watermarking. Meanwhile, it reduced the security breach which exists in reversible watermarking system[6].

This paper proposed a new robust audio zero-watermarking algorithm which has strong robustness toward typical common attacks and hostile attacks. In this algorithm, a binary text image is used as watermarking. The algorithm extracts the low frequency components of host audio to construct zero-watermarking. First of all, the binary text image is scrambled and reduced dimensions to a one-dimensional sequence as watermarking information. This step is to eliminate the correlation of the watermarking images, improves its safety and robustness. Then, it extracts low frequency coefficients of the host audio signals after three-layer wavelet packet decomposition to construct a binary sequence by cubic

spline interpolation. Combining this binary sequence and watermarking information by using an XOR operator, the zero-watermarking is constructed. In the watermarking extraction process, the binary image is reconstructed by using an XOR operator with zero-watermarking and the binary sequence which uses the same method as constructing process to get. At last, copyright authentication is finished by the similarity testing. Throughout the process, the algorithm did not modify the original audio data. That ensures its imperceptibility and achieves blind detection.

2 Theoreticle Basis

2.1 Wavelet Packet Analysis

Wavelet packet transform is a further development based on wavelet transform [7,8,9]. It has been widely used in signal analysis. Fig. 1 is a schematic diagram of three-layer wavelet packet decomposition, where X is the signal, A represents low frequency components, D represents high frequency components, the serial number in the end represents the number of layers of wavelet decomposition.

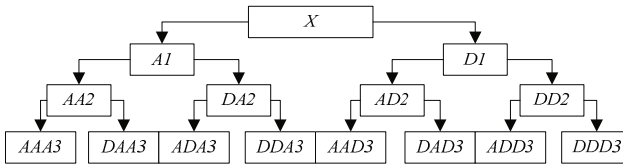


Fig. 1. Schematic diagram of three-layer wavelet packet decomposition of signal X

Low-frequency components of wavelet have strong robustness against various attacks. Therefore, the algorithm uses the first M low-frequency components after wavelet packet decomposition to construct zero-watermarking.

2.2 Cubic Spline Interpolation

Suppose $a < x_0 < x_1 < \dots < x_n < b$ exists on $[a, b]$. If the function meets the demands:

- (1) $s(x) \in C^2[a, b]$;
- (2) $s(x)$ is a cubic polynomial in each little section $[x_i, x_{i+1}] (i = 0, 1, \dots, n-1)$, $s(x)$ is called "Cubic Spline Function" on nodes x_0, x_1, \dots, x_n ; And if the function meets the interpolation demand:
- (3) $s(x) = f_i, i = 0, 1, \dots, n$; $s(x)$ is called "Cubic Spline Interpolation Function" on the interval $[a, b]$.

Section cubic spline interpolation is a smooth curve that goes through the given points. It is concluded from the above three conditions: It is second order

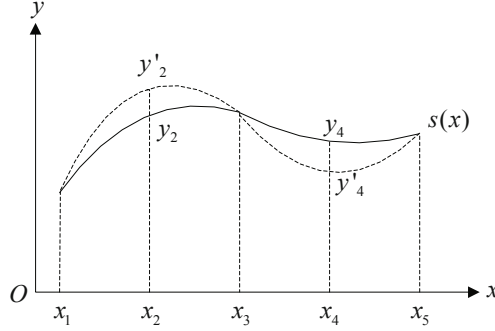


Fig. 2. Cubic spline interpolation schematic

continuous on the interval. It is a cubic polynomial in each little section and goes through the given points. Cubic spline interpolation can be readily finished by the function "interp1(x,y,x_i,'spline')"
in Matlab.

As shown in Fig. 2, according to the given points (x_1, x_3, x_5) on the function $s(x)$, an interpolation function can be obtained by using cubic spline interpolation. The dotted line is drawn to show the interpolation function. Two values y'_2, y'_4 can be obtained by the interpolation points x_2, x_4 . The purpose of getting interpolation is to construct a binary sequence according to comparing the interpolation y'_2, y'_4 and the values of the original curve.

3 Algorithm Description

3.1 Pretreatment of the Watermark Image

Watermark image is a meaningful binary image $V(N * N)$, $V = \{v(i, j), 1 \leq i \leq N, 1 \leq j \leq N\}$, $V(i, j) \in \{0, 1\}$ represents pixel grayscale value of the watermark image matrix(i-th row, j-th column).

Step 1 Matrix V is processed by Arnold scrambling operation.

Step 2 After scrambling operation, watermark image is to be reconstructed into a binary sequence $p(i), i = 1, 2, \dots, N * N$.

3.2 Zero-Watermarking Construction

Fig. 3 is a flowchart of the process of watermarking construction. The original audio processing on the dotted line, under the dotted line is the watermark image preprocessing. Concrete steps are as follows:

Step 1 Decomposes original audio signal $x(n)$ in three-layer wavelet packet[10] and selects 'db4' as the wavelet function. Let $K_2 = N * N$, and extracts the $2K_2 + 1$ -th wavelet low frequency coefficients denoted as $c(i), i = 1, 2, \dots, 2K_2 + 1$. Then mark the odd item $c_{od}(i), i = 1, 2, \dots, K_2 + 1$ and the even item $c_{even}(i), i = 1, 2, \dots, K_2$;

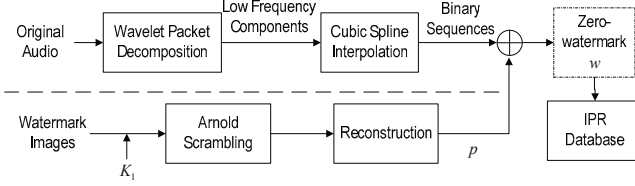


Fig. 3. The flow chart of watermarking construction

Step 2 The one-dimensional array $c_{ood}(i), i = 1, 2, \dots, K_2 + 1$ can be construct into another one-dimensional array measuring $2K_2 + 1$ in length. To extract its even item and mark it $a(i), i = 1, 2, \dots, K_2$;

Step 3 According to the relationship between $a(i)$ and $c_{even}(i)$, to form the binary sequence $u(i)$:

$$u(i) = \begin{cases} 0, & a(i) \geq c_{even}(i) \\ 1, & a(i) < c_{even}(i) \end{cases}, i = 1, 2, \dots, K_2 \quad (1)$$

Step 4 To do XOR operator with $u(i)$ and $p(i)$, then form the zero-watermarking $w(i), i = 1, 2, \dots, k_2$

$$w(i) = p(i) \oplus u(i), i = 1, 2, \dots, k_2. \quad (2)$$

At last, register the IPR information database with the current zero-watermarking, and send and as keys.

3.3 Zero-Watermarking Extraction

Fig. 4 is a flowchart of the process of watermarking extraction. Audio processing is tested on the dotted line, Under the dotted line is watermark image reconstruction and detection process. Concrete steps are as follows:

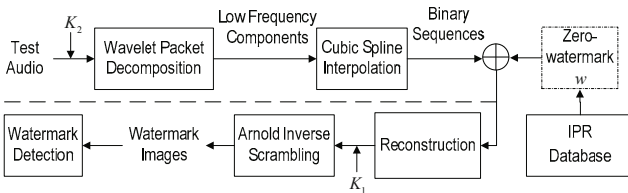


Fig. 4. The flow chart of watermarking extraction

Step 1 Suppose $x'(n)$ is the audio to be checked that attacked in the transmission process. Enter the Key K_2 . and decomposes signal $x'(n)$ in three-layer wavelet packet and selects 'db4' as the wavelet function. Let $K_2 = N * N$, and extracts the $2K_2 + 1$ -th wavelet low frequency coefficients denoted as

$c'(i), i = 1, 2, \dots, 2K_2 + 1$ Then mark the odd item $c'_{ood}(i), i = 1, 2, \dots, K_2 + 1$ and the even item $c'_{even}(i), i = 1, 2, \dots, K_2$;

Step 2 The one-dimensional array $c'_{ood}(i), i = 1, 2, \dots, K_2 + 1$, can be construct into another one-dimensional array measuring $2K_2 + 1$ in length. To extract its even item and mark it $a'(i), i = 1, 2, \dots, K_2$;

Step 3 According to the relationship between $a'(i)$ and $c'_{even}(i)$, to form the binary sequence $u'(i)$:

$$u'(i) = \begin{cases} 0, & a'(i) \geq c_{even}'(i) \\ 1, & a'(i) < c_{even}'(i) \end{cases}, i = 1, 2, \dots, K_2 \quad (3)$$

Step 4 To do XOR operator with $u'(i)$ and the zero-watermarking $w(i)$, then form the zero-watermarking $p'(i), i = 1, 2, \dots, k_2$

$$p'(i) = w(i) \oplus u'(i), i = 1, 2, \dots, k_2. \quad (4)$$

Step 5 Insert the Key K_1 and reform $p'(i)$ as a binary matrix, and do Arnold inverse scrambling with it. The matrix $V' = \{v'(i, j), 1 \leq i, j \leq N\}$ (size $N \times N$) is the watermarking image extracted.

4 Experimental Results and Performance Analysis

Experiment selects Matlab7.8 as simulation software, a digital audio signal (44.1kHz, 16bits) as original audio and a binary image (size 6464) as watermarking image. Take $K_1 = 10$, as shown below:



Fig. 5. Watermarking image

Experiment compares the original watermarking image with the extracted watermarking image by calculating the Normalized Correlation Coefficient (NC). The formula as follows:

$$NC(W, W') = \frac{\sum_i w(i)w'(i)}{\sqrt{\sum_i w^2(i)} \sqrt{\sum_i w'^2(i)}} \quad (5)$$

Where W is original watermarking, W' is the extracted watermarking.

4.1 Common Attacks Testing

To do common attacks with the original audio: 1) Adding white Gaussian noise (mean 0, variance 0.01 or 0.02); 2) Filtering with 6-tap Butterworth filter (cutoff frequency 15 kHz); 3) Re-quantification (32 bits or 8 bits); 4) MP3 compression (128 Kbps or 64 Kbps). The experimental results (NC and extracted watermarking image)algorithm in literature[11]results are shown in Table 1.

Table 1. Common attack experimental results

	Algorithm in literature[12]		This algorithm
	NC	NC	Extracted watermark
Without attack	1	1	水音 印频
Add noise(0.01)	0.9676	0.9904	水音 印频
Add noise(0.02)	0.8781	0.9840	水音 印频
Filtering (15kHz)	0.8955	0.9849	水音 印频
Re-quantification 16816bits	0.9978	0.991	水音 印频
Re-quantification 163216bits	1	0.9991	水音 印频
MP3 compression 128kbps	0.9944	0.9935	水音 印频
MP3 compression 64kbps	0.9242	0.9818	水音 印频

Experimental data show that the algorithm for a typical common attack are reflected robustness. In addition, a number of other common attacks(such as amplitude zoom in, zoom, etc.) anti-attack performance experiments, the results also improved. From the point of view watermarked image: After the audio

being common attack, the extracted watermark image is still clearly visible and uniform distribution of pixels, missing piece of the image did not occur, does not affect the identification of audio watermarking. Proved relatively stable algorithm for common attack, enabling audio copyright authentication. Further, the results from the comparison with the literature[12] can be seen, for most conventional attack, the present algorithm has increased the robustness, particularly because of the low complexity of the algorithm, the program running time is also very improvements.

4.2 Hostile Attacks Testing

Original audio is attacked by the two most typical hostile attacks[12] (cutting and replacement attack). Then the experiment tests the robustness of the algorithm against hostile attacks, according to similarity (NC) and the extracted watermark image.

(1) Cutting Attack

Cut 1/16 length (0-10000 sampling points) of the original audio, the test audio is generated. The extracted watermarking image is shown in Fig. 6 (NC=0.8858).



Fig. 6. Experimental results of cutting attack

(2) Substitution Attack Replace 1/16 length (0-10000 sampling points) of the original audio by a audio segment of the same sampling rate and quantization precision, the test audio is generated. The extracted watermarking image is shown in Fig. 7 (NC=0.8951).



Fig. 7. Experimental results of substitution attack

The experimental results show that when the original audio is processed by hostile attacks, there is a wide gap between the extracted watermarking image and the original. But the image is not one-piece missing. Watermarking identification will not be affected. This proves that algorithm has strong robustness against typical hostile attacks, and can achieve the certification of the audio copyright.

5 Conclusion

This paper proposed a new robust audio zero-watermarking algorithm which can be used to authenticate the copyright of digital audio. It extracts the low frequency components of original audio to construct zero-watermarking by wavelet packet decomposition and cubic spline interpolation. Combining with a binary text images, a meaningful zero-watermarking is constructed and extracted. That completes copyright authentication of the original audio copyright. The zero-watermarking detection process is intuitive, accurate, and with good security. The experimental result shows that this algorithm has strong robustness against typical common attacks and hostile attacks. The algorithm is simple, easy to implement, so it has a high application value.

Acknowledgments. This research supported by the National Nature Science Foundation of China (No.61072087, No.61371193).

References

1. Wu, S., Huang, J., Huang, D.: Efficiently Self-Synchronized Audio Watermarking for Assured Audio Data Transmission. *IEEE Transactions on Broadcasting* 51(1), 69–76 (2005)
2. Nutzinger, M.: Real-time Attacks on Audio Steganography. *Journal of Information Hiding and Multimedia Signal Processing* 3(1), 47–65 (2012)
3. Lahouari, G., Ahmed, B., Mohammad, K.I.: Digital Image Watermarking using Balanced Multiwavelets. *IEEE Transactions on Signal Processing* 54(4), 1519–1536 (2006)
4. Yang, Y., Niu, X.: Multimedia Information Pretend Summarization. *Journal of China Institute of Communication* 23(5), 32–38 (2002)
5. Wen, Q., Sun, T., Wang, S.: Concept and Application of Zero-Watermark. *Acta Electronica Sinica* 31(2), 214–216 (2003)
6. Weng, S., Chu, S., Cai, N., Zhan, R.: Invariability of Mean Value Based Reversible Watermarking. *Journal of Information Hiding and Multimedia Signal Processing* 4(2), 90–98 (2013)
7. Paquet, A., Ward, R., Pitas, I.: Wavelet Packets-based Digital Watermarking for Image Verification and Authentication. *Signal Processing* 83(10), 2117–2132 (2003)
8. Gao, Z., Hai, X.: *Matlab Wavelet Analysis and Application*. National Defense Industry Press, Beijing (2007)
9. Zhang, D.: *Matlab Wavelet Analysis*. China Machine Press, Beijing (2009)
10. Zhong, X., Tang, X.: A DWT Domain Zero-watermark Algorithm Based on Audio's Character. *Journal of Hangzhou Dianzi University* 27(2), 33–36 (2007)
11. Yang, J., Ma, Z., Zhang, X.: Digital Audio Dual Watermarking Scheme Based on Wavelet Packet Analysis. *Journal of Computer Applications* 30(5), 1218–1220 (2010)
12. Steinebach, M., Petitcolas, F., Raynal, F., Dittmann, J., Fontaine, C., Seibel, S., Fates, N., Ferri, L.C.: Stirmark Benchmark: Audio Watermarking Attacks. In: *Int. Conference on Information Technology: Coding and Computing*, pp. 49–54. IEEE Press (2001)

Information Security Management for Higher Education Institutions

Simon K.S. Cheung

The Open University of Hong Kong
Good Shepherd Street, Homantin, Kowloon, Hong Kong
kscheung@ouhk.edu.hk

Abstract. Information security aims at protecting the information assets of an organization from any unauthorized access, disclosure and destruction. For information security to be effectively enforced, good management practices comprising policies and controls should be established. This paper investigates the information security management for higher education institutions. Based on the conventional CIA (confidentiality, integrity and availability) triad of information, eight control areas on information security are identified. They include information asset controls, personnel controls, physical controls, access controls, communication controls, operation controls, information system controls, and incident management and business continuity. A governance framework is important for establishing the policies and executing the controls of information security. It is necessary to maintain a right balance between the technical feasibility and the flexibility and efficiency in administration.

Keywords: information security management, information security policies, information security controls.

1 Introduction

Nowadays, computer systems are highly connected through the internal networks (Intranet) and external networks (Internet) to facilitate accesses to information. This however creates the issue of information security – the protection of information assets from any unauthorized access, disclosure, modification and destruction, in order to ensure its confidentiality, integrity and availability [1, 2]. Information security is conventionally defined as the assurance of the CIA triad of information (confidentiality, integrity and availability) [1, 3], and its extension (authenticity, non-repudiation and accountability) [4, 5].

With the recent advances of communication and mobile technologies, Internet and Intranet accesses to information from client-end devices (especially mobile devices) via wired or wireless networks are very popular, such as on e-mail communication, and e-commerce and e-government services. This inevitably adds more technical complexity in ensuring that the information assets of an organization can be well protected [6, 7, 8], and therefore, some intelligent and sophisticated access protocols are developed [9, 10, 11, 12].

Although there are many technical solutions to help protect the information assets of an organization, the risk of information leakage, modification or destruction cannot be completely eliminated. As this may incur great losses, information security is essential to any organization which counts information assets as critical to their business operation. This is especially important for government and public bodies because the adverse impacts are much greater than that of other organizations [7, 8, 13]. Similarly, for a higher education institution where a large amount of student information is hosted student administrative systems, learning management systems and platforms [14, 15, 16], any information leakage or loss would have large impacts. Information security compliance and awareness have become emerging issues in higher education institutions [17, 18, 19].

In many countries, there are laws, regulations and policies, governing information security, such as the Data Protection Act and Computer Misuse Act in the United Kingdom and the Federal Information Security Management Act in the United States. In Hong Kong, a set of baseline policies have been established for enforcing information security in government offices [20]. Besides, many national and international standards for information security management have been established. Among these standards, the ISO 27001 Information Security Management System is the most widely adopted one [21].

This paper investigates the information security management for higher education institutions. Eight control areas for providing the rules of governance and control of information security are identified, and a framework for governance and control is discussed. These control areas include information asset controls, personnel controls, physical controls, access controls, communication controls, operation controls, information system controls, and incident management and business continuity. The rest of this paper is organized as follows. Section 2 states the principles of information security. Section 3 elaborates the eight key control areas on information security for higher education institutions. Section 4 then discusses the governance of information security. Section 5 briefly concludes this paper.

2 Principles of Information Security

Conventionally, the CIA triad (confidentiality, integrity and availability) forms the principles of information security [1, 3]. In the literature, it has been argued that the CIA triad should be extended with three more principles, namely, authenticity, non-repudiation and accountability [4, 5]. Figure 1 shows these principles.

Confidentiality is the ability to protect information from unauthorized accesses. A typical example of unauthorized accesses is the use of another person's account and password to access an online banking system, which he or she does not possess the necessary access rights. Integrity is the ability to protect information from undetected modification or deletion. For example, in an e-mail communication, some information in the e-mail message is intercepted, modified or omitted during the message sending process. Availability is the ability to protect information from attacks denying or inconveniencing authorized accesses. It ensures that information is readily accessible to the authorized users at all times.

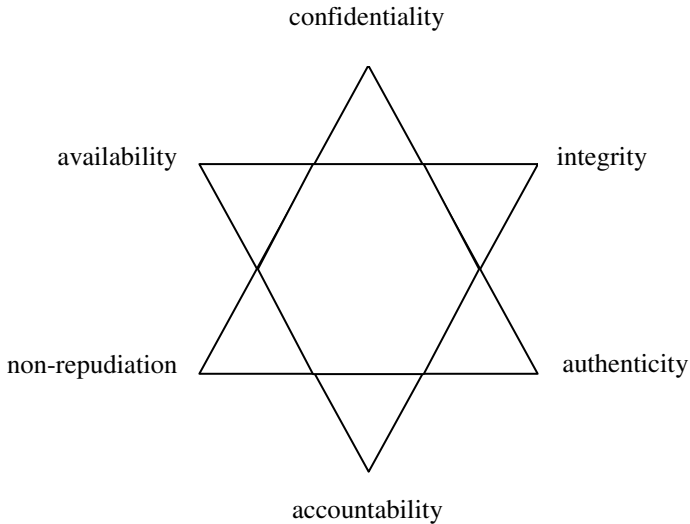


Fig. 1. Six principles of information security

Authenticity is the ability to ensure that transactions or communications of information are genuine. In order to validate accesses to information, authentication system with proper access control and password protection is adopted. Non-repudiation refers to one's intention to fulfill the accepted obligations. For example, in the transmission of information, the sender cannot deny having sent the information and the receiver cannot deny having received the information. Digital signature is often used to ensure non-repudiation. Accountability is the ability to track user identity and actions applied to the information. Accountability is a useful element for executing non-repudiation that proves the performance of an action, for example, sending or receiving information, and when and where the action was performed.

3 Information Security Controls

It is necessary for a higher education institution to establish policies and control measures for ensuring information security [17, 18]. These essentially transform the principles of information security to implementation.

The ISO 27001 Information Security Management System provides a thorough coverage of the key control areas of information security [21]. By making reference to ISO 27001, there are at least eight control areas for a higher education institution, namely, information asset controls, personnel controls, physical controls, access controls, communication controls, operation controls, information system controls, and incident management and business continuity.

3.1 Information Asset Controls

Policies should be established to ensure that appropriate levels of protection and accountability are maintained for information assets. This should be made in accordance with the sensitivity, criticality and values of information assets, regardless of the media on which they are stored, the manual or automated systems that process them, and the methods by which they are distributed. In a higher education institution, information assets should be classified, and the owner, custodian and users of the information assets should be well defined. It is a good practice that an institution should maintain a master record control table which shows a full list of information assets and the owner, custodian and users of the information assets. This control table is referenced in implementing control measures.

3.2 Personnel Controls

Policies should be established to ensure that everyone in an organization clearly understand his or her roles and responsibilities to reduce the risk of theft, fraud or misuse of information assets. For a higher education institution, all staff should be aware of the information security threats and concerns, and are equipped to support information security in the course of their normal work and reduce the risk of human errors. For example, staff in the Registry and Student Affair Office used to handle a large amount of student information. They have the responsibilities of protecting the student information from theft, fraud and misuse. Control measures should be in place to reduce the risk of theft, fraud or misuse of student information. In many institutions, downloading of student information to portable storage is prohibited, unless absolutely required.

3.3 Physical Controls

Policies should be established to ensure that appropriate physical security and control should be maintained to protect against any unauthorized accesses to some defined secure areas such as data centres. For a higher education institution, computer systems and storage of critical and sensitive information shall be housed in data centres with proper physical access controls. Only authorized persons are allowed to have physical accesses to the data centres, and the access logs should be maintained. Besides, proper environment controls should be in place to protect the computer systems and storage from physical damage. Temperature and humidity should be kept at an acceptable level. Gas-based fire extinguishing systems, instead of water-based fire extinguishing systems, should be installed in data centres to minimize the risk of physical damage to data storage devices in case of fire.

3.4 Access Controls

Policies should be established to ensure that access control to information systems and information processing facilities, and that access rights are properly authorized,

allocated and maintained. Control measures should be implemented to enforce authorized accesses to information as well as to reduce the risks of unauthorized access, loss or damage to information. These measures should also be applied to mobile and remote accesses. In a higher education institution, there should be proper access controls for information systems and information processing facilities, where student information and financial information are stored. An access control table should be defined for each information system. Besides, password controls should be enforced, for example, adoption of strong passwords and compulsory changes of passwords over a certain time period.

3.5 Communication Controls

Policies should be established to define procedures for the management and operation of network and communication facilities. Control measures should be implemented to maintain the confidentiality, integrity and availability of communication facilities, such as electronic mailing systems and network storage for information exchange. For a higher education institution, electronic communication is very common. Electronic mails containing student information or sensitive information should be handled with care. It is a good practice to use secured electronic mail systems to protect sensitive information from undetected interception, modification or omission. Encryption and password protection should also be applied to data files on network storage as well as mobile and portable storage devices.

3.6 Operation Controls

Policies should be established to define procedures for the management and operation of computer systems and information processing facilities. Control measures should be implemented to maintain the confidentiality, integrity and availability of the computer systems and information processing facilities. System fixes and patches, especially those related to information security, should be timely applied. Backup procedures should be tightly followed, and tapes and disks should be properly stored. It is a good practice to arrange regular system drills to ensure that all critical systems and facilities can be correctly restored in case of information security incidents. System administrator passwords should be properly maintained, and strict password controls, such as the use of strong passwords and compulsory periodical password changes, should be enforced.

3.7 Information System Controls

Policies should be established to ensure proper controls to prevent information systems from any unauthorized modification and misuse of information. Information security requirements should be clearly identified at the beginning of system development. For a higher education institution, the input, processing and output of student information should be properly defined and implemented. These should be enforced during the acquisition, development and maintenance of information

systems. All changes on information systems should be logged. It is a good practice that regular review on these information systems should be conducted to identify and fix information security loopholes if any.

3.8 Incident Management and Business Continuity

Policies should be established to ensure that information security incidents are communicated in an appropriate manner, allowing timely corrective actions to be taken. Clear procedures should be set out for handling incidents that might have an impact on information security. Incidents should be classified in term of severity and impact. According to the level of severity and the scope and impact of an incident, an appropriate incident coordinator should be appointed. On the other hand, it is a good practice that critical business processes identified and integrated with information security requirements, in order to minimize the impact to an acceptable level. For a higher education institution, teaching and learning are critical, and hence, control measures should be enforced to maintain continuity of teaching and learning activities in case of information security incidents.

4 Governance of Information Security

A governance framework is important for establishing the policies and executing the controls of information security. This section discusses the governance of information security in a higher education institution.

It is a good practice to appoint an information security officer who is responsible for the overall governance of information security. In practice, there are two models for information security governance, namely, executive-led model and committee-led model. In the executive-led model, the information security officer is a senior officer who takes the overall responsibility of information security for the institution, including decision-making and policy-making. In the committee-led model, an information security committee is established to take up the roles of an information security officer. Chaired by a senior officer, the committee comprises the owners and custodian of major information repositories, such as the Registrar, Secretary, and the Director of Information Technology.

The information security officer or information security committee is wholly responsible for the design, implementation and execution of the policies and measures on information asset controls, personnel controls, physical controls, access controls, communication controls, operation controls, information system controls, and incident management and business continuity. It is important that appropriate authority should be given to the information security officer or information security committee for discharging these duties and responsibilities, especially in handling information security incidents and problems.

Besides implementing the policies and executing the controls, the information security officer or information security committee should conduct regular review on the compliance of information security. A typical way to review the compliance is to

conduct an information security audit. Like many security audits, an information security audit aims to check the compliance of information security with respect to the established policies, guidelines and procedures [22, 23, 24, 25]. It is a good practice for a higher education institution to establish its own audit schedule on information security. Some well-known standards can be referenced in establishing an information security audit framework [24, 25].

Finally, a higher education institution should always ensure that all its staff and students have the awareness on information security, and a thorough understanding of the prevailing policies and controls of information security. To serve this purpose, regular trainings and briefing sessions on information security should be conducted. They are especially useful for new staff and new students, and therefore better be held at the start of each semester. In addition to these trainings and briefing sessions, from time to time, any updates on the information security policies and controls should be communicated to all staff and students.

5 Conclusion

Information security is essential to higher education institutions as any information leakage and damage would incur great losses. There is a need for a higher education institution to enforce information security. Based on the principles of information security, we identify eight control areas on information security, namely, information asset controls, personnel controls, physical controls, access controls, communication controls, operation controls, information system controls, and incident management and business continuity. Policies, guidelines and control measures should be established. While the policies provide rules of governance of information security, the guidelines and control measures help execute and implement the policies.

It is important to address a salient point in establishing the policies, guidelines and controls on information security. In reality, flexibility and control are contradictory to each other. In order to enforce information security, it is necessary to implement control measures which inevitably create inflexibility and inconvenience. A right balance between flexibility and control is however difficult to achieve. There are also administrative considerations in implementing the policies, guidelines and control measures, such as on the availability of resources and efficiency in administration. A strong support from senior management is absolutely necessary.

References

- [1] Bishop, M.: Computer Security, Art and Science. Addison-Wesley (2003)
- [2] Raggad, B.G.: Information Security Management: Concepts and Practices. CRC Press (2010)
- [3] Peltier, T.: Information Security Policies and Procedures: A Practitioner's Reference. CRC Press (2004)

- [4] Parker, D.B.: Toward a New Framework for Information Security. In: Kabay, M.E. (ed.) *The Computer Security Handbook*. John Wiley (2002)
- [5] Anderson, J.M.: Why We Need a New Definition of Information Security. *Computer and Security* 22(4), 308–313 (2003)
- [6] Matbouli, H., Gao, Q.: An Overview on Web Security Threats and Impact to e-Commerce Success. In: *Proceedings of the International Conference on Information Technology and e-Services*, pp. 1–6. IEEE Press (2012)
- [7] Singh, S., Karaulia, D.S.: E-Governance: Information Security Issues. In: *Proceedings of the International Conference on Computer Science and Information Technology*, pp. 120–124. IEEE Press (2011)
- [8] Hwang, M.S., Li, C.T., Shen, J.J., Chu, Y.P.: Challenges in e-Government and Security of Information. *Information & Security* 15(1), 9–20 (2004)
- [9] Akhawe, D., Barth, A., Lam, P.E., Mitchell, J.: Towards a Formal Foundation of Web Security. In: *Proceedings of the IEEE Symposium on Computer Security Foundations*, pp. 290–304. IEEE Press (2010)
- [10] Pansa, D., Chomsiri, T.: Web Security Improvement by using Dynamic Password Authentication. In: *Proceedings of the International Conference on Network and Electronic Engineering*, pp. 32–36. IACSIT Press (2011)
- [11] Chen, C.M., Wang, K.H., Wu, T.Y., Pan, J.S., Sun, H.M.: A Scalable Transitive Human-Verifiable Authentication Protocol for Mobile Devices. *IEEE Transactions on Information Forensics and Security* 8(8), 1318–1330 (2013)
- [12] Chen, C.M., Chen, Y.H., Lin, Y.H., Sun, H.M.: Eliminating Rouge Femtocells based on Distance Bounding Protocol and Geographic Information. *Expert Systems with Applications* 41(2), 426–433 (2014)
- [13] Cheung, K.S.: Development of Organizational Information Security Policies. In: *Proceedings of the International Conference on Intelligent Computing and Intelligent Systems*, pp. 753–756. IEEE Press (2011)
- [14] Cheung, K.S.: A Comparison of WebCT, Blackboard and Moodle for the Teaching and Learning of Continuing Education Courses. In: Tsang, P., et al. (eds.) *Enhancing Learning Through Technology*, pp. 219–228. World Scientific (2006)
- [15] Yau, J., Lam, J., Cheung, K.S.: A Review of E-Learning Platforms in the Age of E-Learning 2.0. In: Wang, F.L., Fong, J., Zhang, L., Lee, V.S.K. (eds.) *ICHL 2009*. LNCS, vol. 5685, pp. 208–217. Springer, Heidelberg (2009)
- [16] Cheung, K.S., Lam, J., Yau, J.: A Review of Functional Features of E-Learning Platform in the Continuing Education Context. *International Journal of Continuing Education and Lifelong Learning* 2(1), 103–116 (2009)
- [17] Rezgui, Y., Marks, A.: Information Security Awareness in Higher Education: An Exploratory Study. *Computers & Security* 27(7), 241–253 (2008)
- [18] Kvakik, R.B.: *Information Technology Security: Governance, Strategy and Practice in Higher Education*, Center for Applied Research, EDUCAUSE (2004)
- [19] Kam, H.J., Katerattanakul, P., Gogolin, G., Hong, S.: Information Security Policy Compliance in Higher Education: A Neo-Institutional Perspective. In: *Proceedings of the Pacific Asia Conference on Information Systems*. Association for Information Systems (2013)
- [20] OGCIO, Baseline IT Security Policy, The Office of the Government Chief Information Officer, The Government of the Hong Kong Special Administrative Region, Hong Kong (2009)

- [21] ISO, ISO 27000 : Information Security Management System : Family of Standards, Joint Technical Committee, International Organization for Standardization and International Electrotechnical Commission (2005)
- [22] Onwubiko, C.: A Security Audit Framework for Security Management in the Enterprise. In: Jahankhani, H., Hessami, A.G., Hsu, F. (eds.) ICGS3 2009. CCIS, vol. 45, pp. 9–17. Springer, Heidelberg (2009)
- [23] Lo, E.C., Marchand, M.: Security Audit: A Case Study. In: Proceedings of the Canadian Conference on Electrical and Computer Engineering, pp. 193–196. IEEE Press (2004)
- [24] Kelson, N.: Information Security Management Audit and Assurance Programme. In: ISACA (2010)
- [25] ISO, ISO 27007 : Guidelines for Information Security Management Systems Auditing, Joint Technical Committee, International Organization for Standardization and International Electrotechnical Commission (2011)

Laser Induced Breakdown Spectroscopy Data Processing Method Based on Wavelet Analysis

Lu Muchao

Taiyuan University of Technology College of Information Engineering,
Taiyuan 030024, China

Abstract. In this paper, we present a data processing approach for Laser induced breakdown spectroscopy (LIBS). This method is based on wavelet analysis and pattern matching. First, it uses wavelet transforms to decompose the laser induced spectrum data which comes from the sample and obtain the decomposition coefficient of spectrum, then reconstructs the feature background spectrum by means of low frequency coefficient. Through using pattern cluster method to divide the spectrum data of calibration sample into some subsets, then do the calibration for each spectra data in each subsets. Second, we extract effective measurement pattern class template and calibration parameter from the spectrum subset which has the minimum differ between the result of calibration sample and the reality value. In practical process of measurement, we use effective measurement pattern class template to match the spectra data to identify the effectiveness of the measurement. Therefore, we can calculate element contents with the calibration parameter achieved before. This method can decrease the times of laser excitation and increase the measurement accuracy effectively.

1 Introduction

Laser Induced Breakdown Spectroscopy (LIBS) is an analysis technique, in which spectra of laser-produced plasmas were used for qualitative as well as quantitative spectrochemical analysis of material[1]. During the past decade, related technology has produced more reliable lasers, charge coupled detectors, and miniature spectrographs with its capabilities of recording spectra over a wide range of wavelengths. The combination of these technologies has produced unprecedented enhancements in the signal-to-noise ratio. LIBS has rapidly developed into a major analytical technology with the capability of detecting all chemical elements in a sample without any preparation, of real-time response, and of close-contact or stand-off analysis of targets. So it will be used widespread in the future.

But since the spectrum plasma may be disturbed by matrix effect and some objective factors which are difficult to avoid, such as laser intensity fluctuation, characteristics of the sample surface, laser-induced breakdown spectroscopy has some problems which are large random and poor reproducibility and it influence the accuracy of the quantitative analysis.

This paper uses wavelet transform method to obtain the effective measurement pattern class template. Then using this template to match the measured spectra

data, identify the validity of the spectrum and calculate the element contents of material. It can increase the accuracy of measurement and reduce the stimulating times of laser.

2 Measuring Principle of Laser-Induced Spectroscopy

LIBS uses a beam of intense pulsed laser irradiation to the measuring material after focusing, the focal point of the measurement object ionization and generate high temperature, high-density plasma. In this method, a solid target is vaporized by a powerful laser pulse to form partially ionized plasma that contains atoms and small molecules. In the high temperature system, the fierce collision between the particles makes the molecular or atomic ionized into ions, and the molecular, atomic and ions can distribute on all energy level, high energy level transition to low level so as to make the laser plasma generate strong spectrum. The LIBS system diagram is shown in the Fig. 1.

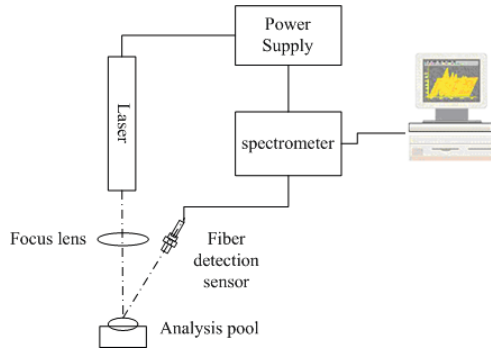


Fig. 1. LIBS system diagram

If local thermodynamic equilibrium (LTE) condition is assumed, the re-absorption effects are negligible (i.e. the plasma is optically thin)[2], the spectrally integrated line intensity, corresponding to the transition between levels E_k and E_i of the generic atomic species with concentration N_s , can be expressed as

$$I_{\lambda}^{K_i} = N_s A_{ki} \frac{g_k e^{-(E_i/K_a T)}}{U_s(T)} \quad (1)$$

where λ is the transition wavelength, N_s is the density of emission atomic, A_{ki} is the transition probability of this spectral line, $U_s(T)$ is the partition function under the plasma temperature, the intensity unit of the line of departure is photon number/ cm^3 , in the actual measurement process, take the efficiency of optical receiver system into consideration, the intensity of experimental spectral line is indicated as[3]

$$\overline{I_{\lambda}^{K_i}} = F C_s A_{ki} \frac{g_k e^{-(E_i/K_a T)}}{U_s(T)} \quad (2)$$

$\overline{I_{\lambda}^{K_i}}$ is the line intensity of the measurement, C_s is the atomic content correspond to this emission line, F is the experimental system parameters including optics efficiency of the receiving system and plasma temperature and its volume. In order to intensify the intensity of spectrum signal and increase the SNR, it calculates the average value of multi-spectrum. The formula (2) only contains C_s as unknown parameters and C_s is related to the element contents of the measured material, the other parameters are known. So when obtaining the $\overline{I_{\lambda}^{K_i}}$, the intensity of this correspond spectral line expresses the concentration of the analysis element by using calibration.

LIBS requires the measuring system under a strict stable environment, such as the laser energetic and point focusing must stand stable[4]. Unfortunately the system cannot keep stable under numbers of conditions in the actual measurement, so it will make a difference between every stimulate results and the simple average method cannot make the efficient handling.

3 The Method Based on Wavelet Feature Extraction

Because the laser induced spectrum contains many spectrum lines, so it is hard to classify and identify the effective or not. In order to solve this problem, we use wavelet decomposition and reconstruction to obtain the background spectrum. After that we classify the background spectrum and obtain the effective measurement pattern, which is used as a template to judge effective of measurement data. And next, we extract the calibration parameters which come from the effective schema. In the actual measuring process, we use the template as the criteria to determine each measurement result whether it is effective or not. If the data set is effective, then calculate the element content via calibration parameters. It is important that this method can obtain effective measuring data in a few excitation processes and increase the measuring accuracy. The data handle process is shown as Fig. 2.

3.1 Wavelet Decomposition and Reconstruction

Recently, the wavelet transform has received considerable attention from researchers in many areas such as signal processing, image processing, pattern recognition, communication, etc. The primary attractive feature of wavelet transform is its capacity for multiresolution analysis. It achieves low-frequency resolution and high time resolution in the high-frequency band, and high-frequency resolution and low time resolution in the low-frequency bands in an adaptive manner. at the same time, Wavelet transform (WT) exhibits very attractive features that make it ideal for studying spectrum signals, so in recent year, wavelet has been used widely in spectrum data processing[6][7][8].

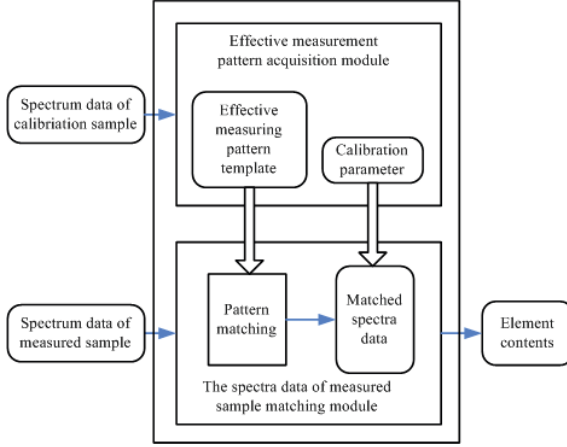


Fig. 2. Diagram of LIBS spectra data processing

Wavelet transform decomposes a signal into localized contributions labeled by a scale and a position parameter. And each of the contributions at different scale represents the information of different frequency contained in the original signal. The discrete wavelet transform(DWT) decomposes the time record $x(t)$ ($t=1,2, \dots, N$) into dyadic wavelet functions $\psi_{j,k}(t)$ and scaling functions $\varphi_{j,k}(t)$. The basis for this decomposition is formed from mother wavelet $\psi(t)$ and father wavelet $\varphi(t)$, by translating in time and dilating in scale[5].

$$\begin{aligned} \psi_{j,k}(t) &= 2^{-j/2} \psi(2^{-j}t - k) \\ \varphi_{j,k}(t) &= 2^{-j/2} \varphi(2^{-j}t - k), \quad j, k \in Z \end{aligned} \quad (3)$$

where $k=1, 2, \dots, N/2$, N is the length of data queue. $j=1, 2, \dots, J$, J is often a natural number, Z is the set of integers. Wavelet decomposition produces a family of hierarchically organized decompositions.

At each level j , the j -level approximation $A_j(t)$, and a deviation signal called the j -level detail $D_j(t)$ can be calculated according to the following equations.

$$D_j(t) = \sum_{k \in Z} W(j, k) \psi_{j,k}(t) \quad j, k \in Z \quad (4)$$

where, $W(j, k)$ is the wavelet coefficients, and

$$W(j, k) = \int_{-\infty}^{+\infty} x(t) \psi_{j,k}(t) dt \quad (5)$$

The signal $x(t)$ is the sum of all the details:

$$x(t) = \sum_{j \in Z} D_j(t) \quad (6)$$

Then, take a reference level called J ; there are two sorts of details. Those associated with indices $j \leq J$ correspond to the fine details, the others, which correspond to $j > J$, are the coarser details, we group these latter details into

$$A_J(t) = \sum_{j>J} D_j(t) \quad (7)$$

which defines what is called an approximation of the signal $x(t)$. Apparently, with the increase of the level J , the resolution defined as 2^{-J} decreases, and $A_J(t)$ will only contain the “lower frequency” components of $x(t)$ [5].

3.2 Extraction of Effective Model Class

In experiment, We uses laser to excite m times to each sample which comes from a set of calibration samples of n , every spectra data denoted as $G_{i,j}, i = 1, 2, \dots, n, j = 1, 2, \dots, m$, and it forms the calibration sample measuring spectra data set $\mathbf{G} = \{G_{i,j}\}$. Every spectrum data sequence is expressed as $G_{i,j}(k) = [X_1, X_2, \dots, X_k, \dots, X_N]$, N is the length of the spectrum data sequence. The flowchart of extracting effective pattern is shown in Fig. 3.

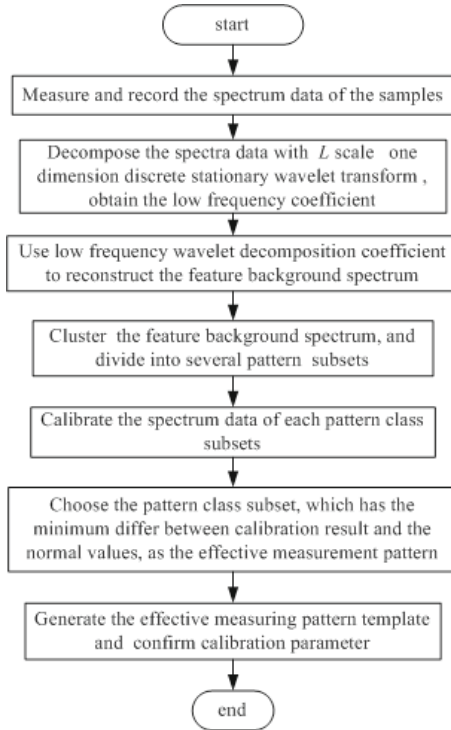


Fig. 3. Flowchart of extracting effective measurement pattern

(1) It decompose each spectrum data sequence $G_{i,j}$, which in the calibration sample spectrum data set, by using L scale one-dimension discrete stationary wavelet and obtains low frequency decomposition coefficient $W_{i,j}^a = [w_{l,k}^a]_{L \times N}$.

(2) Using low frequency decomposition coefficient $W_{i,j}^a$ to reconstruct the spectrum and obtain the feature background spectrum $G_{i,j}^b$, correspond to the spectrum $G_{i,j}$. The feature background data sequence $G_{i,j}^b$ can express as $G_{i,j}^b(k) = [X_1^b, X_2^b, \dots, X_k^b, \dots, X_N^b]$. All of the feature background spectrum $G_{i,j}^b$ compose the feature background spectrum set $\mathbf{G}^b = \{G_{i,j}^b\}$.

(3) Carries on the cluster analysis to the background spectrum data $G_{i,j}^b$ in the feature background spectrum data set \mathbf{G}^b , and dividing the feature background spectrum data set into several pattern class subsets \mathbf{G}_h^b , that is $\mathbf{G}^b = \{\mathbf{G}_1^b, \mathbf{G}_2^b, \dots, \mathbf{G}_h^b, \dots, \mathbf{G}_H^b\}$, $h = 1, 2, \dots, H$. Here H is the number of pattern class subsets which are obtained by analyzing the feature background spectrum data set. According to the correspondence relationship between the spectrum measuring data $G_{i,j}$ and the feature background spectrum data $G_{i,j}^b$, and the partition of background spectrum data set $\mathbf{G}^b = \{G_{i,j}^b\}$, we can divide the calibration sample spectrum measuring data set \mathbf{G} into several pattern class subsets \mathbf{G}_h which are correspond to the pattern class subset \mathbf{G}_h^b of feature background data set \mathbf{G}^b , that is $\mathbf{G} = \{\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_h, \dots, \mathbf{G}_H\}$.

(4) Using reality element content to calibrate the spectrum data contained in each subset \mathbf{G}_h , each subset \mathbf{G}_h can obtain a set of calibration parameter β_h and calibration calculating result. Choosing the subset which has the minimum differ between the calibration calculation result and reality value, then extract the feature parameters of this pattern class to form effective measuring pattern G_m . The method of extract effective measuring pattern G_m as follows:

Suppose the subset \mathbf{G}_h has the minimum differ between the calculated result and reality value. The calibration sample measuring data subset \mathbf{G}_h correspond to the feature background spectrum subset \mathbf{G}_h^b which has E numbers feature background spectrum and the sequence length of each feature background spectrum data is N . Choosing the maximum value of the k locations among all the spectrum sequence of \mathbf{G}_h^b as the higher limit of the effective measuring pattern class sequence $G_h^m(k)$, and choosing the minimum value of the k locations among all the spectrum sequence of \mathbf{G}_h^b as the lower limit of the effective measuring pattern class sequence $G_l^m(k)$, programming language described by the following:

for $k = 1$ to N

$$G_h^m(k) = \max_{i=1}^E(G_i^b(k)); \quad G_l^m(k) = \min_{i=1}^E(G_i^b(k));$$

end

The effective measuring pattern class model defined as $G_m = [G_l^m(k), G_h^m(k)]$.

(5) Choosing the calibration results of $\mathbf{G} = \{\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_h, \dots, \mathbf{G}_H\}$ and the calibration parameter β_h (correspond to the subset \mathbf{G}_h which has minimum differ between the results and calibration sample element content reality value) as the calibration parameter used in the actual measuring, and involved in calculating the measured sample element content.

3.3 The Application of Effective Pattern Template

Excite the actual sample to obtain the single laser induced spectrum data G_j , $j = 1, 2, \dots$.

(1) Process L scale one-dimension discrete stationary wavelet transform to decompose the laser induced spectrum data and obtain the low frequency decomposition coefficient $W_{i,j}^a = [w_{l,k}^a]_{L \times N}$.

(2) Use low frequency wavelet decomposition coefficients W_j^a to reconstruct the spectrum and obtain the feature background spectrum G_j^b which is correspond to the spectrum G_j .

(3) Use effective measuring pattern class template to match the feature background spectrum G_j^b , if this feature background spectrum G_j^b belongs to the effective measuring pattern class, this measured spectrum data is regarded as effective. Feature background spectrum G_j^b and the effective measuring pattern class template matching method is: If the data (location k) in the feature background spectrum data sequence $G_j^b(k)$ fulfill the condition $G_l^m(k) \leq G_j^b(k) \leq G_h^m(k)$, then match the feature background spectrum G_j^b and the effective measuring pattern class template.

(4) Excite the measured sample until the numbers of effective measuring spectrum data beyond the predefined numbers.

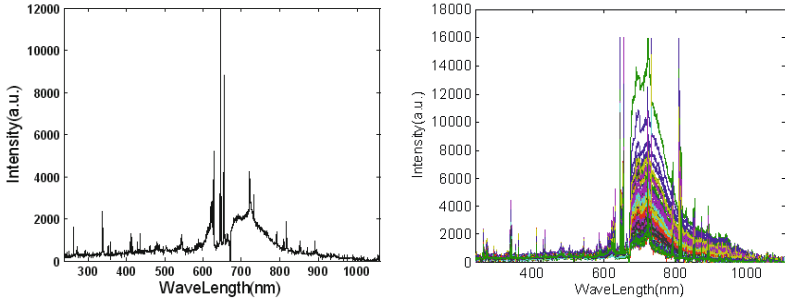
(5) Calculate the element content of the obtained effective measuring spectrum data according to the calibration parameters and use the average value of measured results as the analysis output results.

4 Experiment and Analysis

This paper use LIBS to measure the unburned carbon contents of fly ashes in the thermal power plant. The experiment uses passively Q-type Nd:YAG laser, center wavelength is 1064nm, pulse width is 10ns, pulse repetition frequency is 1~10 Hz, laser energetic are 120~160 mJ/Pulse. The spectrograph is AvaSpec-2048FT, communicate with the computer through the USB interface to transfer spectrum data and receive the control order. The spectrograph sends trigger signal to control laser.

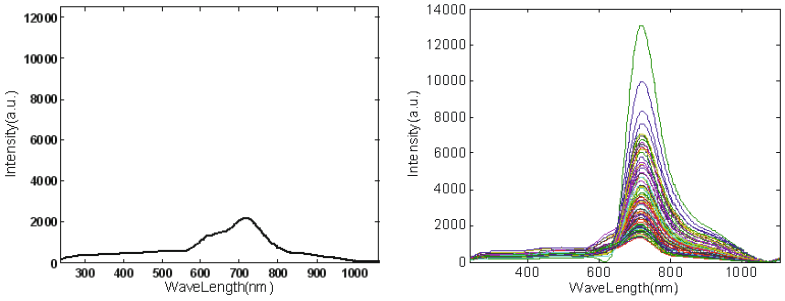
We collect 70 fly ash samples from different regions. Before measure the samples, we grind and stir the samples to make them equality. Then we choose 20 kinds of samples as the calibration samples and performing 100 times laser induced spectrum measuring to each calibration samples and obtain 20×100 laser induced spectrum data. Fig. 4(a) is the spectrum of one time exciting to sample, Fig. 4(b) is the laser induced spectra sets of 100 times exciting to one sample, we can see from the Fig. 4(b) that the random factor and laser energetic fluctuation can affect the stability of spectrum data.

According to this method, a feature background spectrum is reconstructed shown in Fig. 5(a), which corresponding to Fig. 4(a). Fig. 5(b) is the feature background spectrum sets corresponding to Fig. 4(b). Following the next step, the effective measurement pattern template is achieved, which is shown in Fig. 6.



(a) Laser induced spectrum of a single measurement (b) Laser induced spectrum of 100 times measurement to one sample

Fig. 4. Laser induced spectrum of one fly ash sample



(a) Feature background spectrum corresponding to Fig. 4(a) (b) Feature background spectrum corresponding to Fig. 4(b)

Fig. 5. Feature background spectrum of one fly ash sample

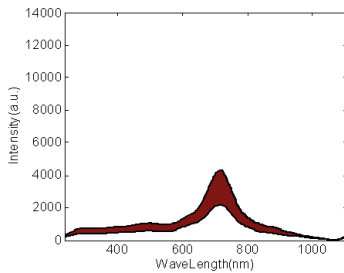


Fig. 6. Sketch map of effective pattern class template

In the measuring process, we use the effective pattern class template to match the measuring data, and if the laser induced spectrum we obtained is effective, then using relevant calibration parameters to calculate the element content. In order to check the effectiveness of this method, we compare the method with the traditional data average method. The traditional method process 50 times

measurement and wipe out 5 maximum values and 5 minimum values, then using the 40 remaining measurement results to calculate the average value, the linear regression result is shown in Fig. 7(a). The method in this paper is using 10 pattern class templates matching effective results to calculate the average, the linear regression result is shown in Fig. 7(b). Comparing Fig. 7(a) and Fig. 7(b), we can clearly figure out that the method of this paper is effective than the average method, also the laser excite times to 50 samples are decrease from 2500 to 956, so it can not only increase the measurement efficiency , but also extend life span of the laser at the same time.

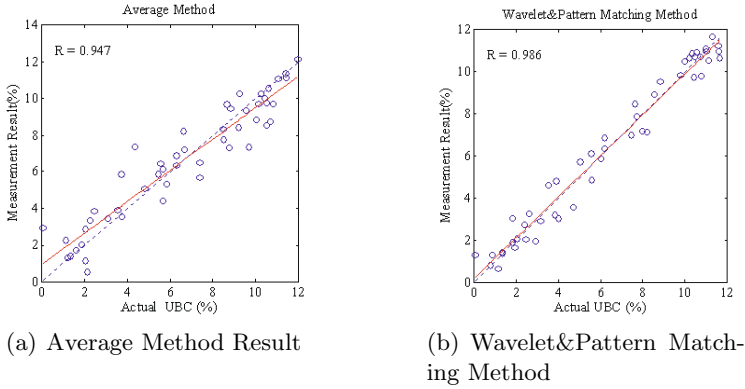


Fig. 7. Result comparison

5 Conclusion

LIBS is widely used in qualitative and quantitative analysis the element contents of various kinds of subjects and LIBS has great practical value. But it has lots of factors which are difficult to avoid, such as laser intensity pulse, feature of sample surface and cardinal effect, so the consequence is that it has large randomness and negative repeatability and the accuracy of quantitative analysis will also be affected. On one hand, we can improve the method by improving the hardware device, on the other hand we can take some proper technique method to process the data of laser induced spectrum. This paper uses wavelet analysis and pattern recognition method to obtain the laser induced spectrum and pick up the effective measuring pattern template and correspond calibration parameter. In measuring process, this paper uses template matching method to identify the effectiveness of measuring data and calculate the effective data. This improvement can increase the measurement effective and accuracy, also can decrease the stimulate times of laser and increase the life span of laser at the same time. The experiment indicate the this method is effective.

References

1. Yao, M., Liu, M., Zhao, J., et al.: Identification of Nutrition Elements in Orange Leaves by Laser Induced Breakdown Spectroscopy. In: Third International Symposium on Intelligent Information Technology and Security Informatics, IITSI 2010, pp. 398–401. IEEE Computer Society, Jingtangshan (2010)
2. Kompitsas, M., Roubani-Kalantzopoulou, F., Bassiatis, I.: Laser Induced Plasma Spectroscopy (Lips) as an Efficient Method for Elemental Analysis of Environmental Samples. In: Proceedings of EARSeL-SIG-Workshop LIDAR, Dresden/FRG, June 16-17, pp. 130–138 (2000)
3. Ciucci, A., et al.: New Procedure for Quantitative Elemental Analysis by Laser-Induced Plasma Spectroscopy. *Appl. Spectrosc.* 53, 960–964 (1999)
4. Ramil, A., Lpez, A.J., Yez, A.: Application of artificial neural networks for the rapid classification of archaeological ceramics by means of laser induced breakdown spectroscopy (LIBS). *Applied Physics A: Materials Science and Processing* 92(1), 197–202 (2008)
5. Hu, Y., Jiang, T., Shen, A., Li, W., Wang, X., Hu, J.: A background elimination method based on wavelet transform for Raman spectra. *Chemometrics and Intelligent Laboratory Systems* 85(1), 94–101 (2007)
6. Esteban-Diez, I., Gonzalez-Saiz, J.M., Gomez-Camara, D., Pizarro Millan, C.: Multivariate calibration of near infrared spectra by orthogonal wavelet correction using a genetic algorithm. *Analytica Chimica Acta* 555, 84–95 (2006)
7. Yiou, P., Sornette, D., Ghil, M.: Data-adaptive wavelets and multi-scale singular-spectrum analysis. *Physica D* 142, 254–290 (2000)
8. Zhang, X., Zheng, J., Gao, H.: Curve fitting using wavelet transform for resolving simulated overlapped spectra. *Analytica Chimica Acta* 443, 117–125 (2001)

Towards Time-Bound Hierarchical Key Management in Cloud Computing

Tsu-Yang Wu^{1,2}, Chengxiang Zhou¹, Eric Ke Wang^{1,2}, Jeng-Shyang Pan^{1,2},
and Chien-Ming Chen^{1,2,*}

¹ Shenzhen Graduate School, Harbin Institute of Technology,
Shenzhen, 518055, China

² Shenzhen Key Laboratory of Internet Information Collaboration,
Shenzhen, 518055, China

{wutsuyang,hitcms2009,jengshyangpan,chienming.taiwan}@gmail.com,
962982698@qq.com

Abstract. Nowadays, data outsourcing in the cloud is used widely and popularly by people. It also arises several security problems. To control access of outsourced data with different priority becomes an important research issue. Recently, Chen et al. proposed the first hierarchical access control scheme in cloud computing. However, they did not concern with the time-bound property. In some applications such as Pay-TV, the time-bound property is necessary because subscriber may subscribe some channels during one month. In this paper, we propose the first time-bound hierarchical key management scheme in cloud computing without tamper-resistant devices. The security analysis demonstrates that the proposed scheme is provably secure against outsider and insider attacks.

Keywords: Time-bound hierarchical key management, cloud computing, bilinear pairing, security.

1 Introduction

Cloud storage services [1] have been received much attention recently. They provide relative techniques for data outsourcing, access [2], and sharing [3]. In data outsourcing, data provider (DP) outsources her/his data to cloud server (CS) rather than storing data locally. Any authorized user can access these data from CS via Internet. However, data outsourcing arises some security problems: (1) DP do not want to disclose her/his data to CS and (2) DP should control access of outsourced data with different priority.

The access control problem is to control users who are able to access the resources in a system. In the system, users may be organized in a hierarchy formed by several disjoint classes. These classes have different limitations on the resources. In other words, some users own more access rights than others.

* Corresponding author.

Up to now, several hierarchical key management schemes have been published in [4,5,6,7]. However, in some applications, time-bound property should be concerned. For example, in Pay-TV system, a subscriber may want to subscribe the news channel in some time period such as one week, one month, or one year. Hence, time-bound property needs involved in hierarchical key management schemes.

In 2002, Tzeng [8] proposed the first time-bound hierarchical key assignment scheme. However, his scheme was proved insecure against the collusion attack Yi and Ye in 2003 [9]. In 2004, Chien [10] proposed an efficient time-bound hierarchical key assignment scheme. Unfortunately, his scheme was also proved insecure in [11]. In 2005, Yeh [12] proposed an RSA-based hierarchical key assignment scheme. In the same year, Wang and Laih [13] proposed a time-bound hierarchical scheme by using merging. In 2006, Ateniese et al. [14] considered the unconditionally secure and computationally secure setting for a time-bound hierarchical scheme. They proved that Yeh's scheme [12] is insecure and proposed a provably-secure time-bound hierarchical key assignment scheme based on the discrete logarithm problem with a tamper-resistant device. In 2009, Sui et al. [15] proposed a time-bound access control scheme for dynamic access hierarchical. This scheme is the first one to support dynamics of access hierarchical. In 2012, Chen et al. [16] proposed a time-bound hierarchical key management scheme without tamper-resistant device. In the same year, Tseng et al. [17] proposed two pairing-based time-bound key management schemes without hierarchy. The first scheme used Lucas function for continuous time period and the second scheme is based on RSA for discrete time period. In 2013, Chen et al. [18] proposed the first hierarchical access control scheme in cloud computing. In their scheme, the encrypted data provided by DP can be transformed by using proxy re-encryption method such that authorized users can decrypt them. Unfortunately, they did not concern the time-bound issue. It means that the user revocation is complex and all established keys must be reset.

In this paper, we based on Chen et al.'s scheme [16] propose the first pairing-based time-bound hierarchical key management scheme in cloud computing. The advantage of our scheme does not require any tamper-resistant devices and is suitable for cloud environments. The security analysis is demonstrated that our scheme is secure against outsider and insider attacks. Finally, the performance analysis is given.

The rest of this paper is organize as follows: In Section 2, we introduce the necessary preliminaries which contain bilinear pairings, the hierarchical access control policy, and all-or-nothing transformation. The concrete scheme is proposed in Section 3. In Section 4, we present the security analysis of our scheme. The performance analysis is given in Section 5 and the conclusions are draw in Section 6.

2 Preliminaries

2.1 Bilinear Pairings

Let G_1 and G_2 be two groups with a same order q , where q is a large prime. Here, G_1 is an additive cyclic group and G_2 is a multiplicative cyclic group. A bilinear pairing e is a map defined by $e : G_1 \times G_1 \rightarrow G_2$ which satisfies the following three properties:

- (1) Bilinear: For all $P, Q \in G_1$, $a, b \in Z_q$, we have $e(aP, bQ) = e(P, Q)^{ab}$.
- (2) Non-degenerate: For all $P \in G_1$, there exists $Q \in G_1$ such that $e(P, Q) = 1_{G_2}$.
- (3) Computable: For all $P, Q \in G_1$, there exists an efficient algorithm to compute $e(P, Q)$.

For the details of bilinear pairings, readers can refer to [19,20,21].

2.2 HAC Policy

The hierarchical access control (HAC) policy enables data access in a hierarchy [22]. According to the HAC policy, data is organized into n classes C_1, C_2, \dots, C_n . The relation between these classes is defined as a binary relation \prec . Note that $C_j \prec C_i$ means that the security level of C_i is higher than C_j . In other words, if a user is allowed to access data in C_i , he can also access data in C_j . However, the opposite is forbidden.

2.3 AONT

An all-or-nothing transformation (AONT) [23] maps an α -blocks message $X = X_1 || X_2 || \dots || X_\alpha$ with a random string r into an α' -blocks message $Y = Y_1 || Y_2 || \dots || Y_{\alpha'}$. AONT satisfies the following three properties:

- (1) $Y \leftarrow AONT(X, r)$ can be computed efficiently for given X and r .
- (2) $X \leftarrow AONT^{-1}(Y)$ can be computed efficiently for given Y .
- (3) It is infeasible to recover X for any block of Y lost.

2.4 Notations

The following notations which are used throughout this paper:

- e : a bilinear map, $e : G_1 \times G_1 \rightarrow G_2$.
- P : a generator of group G_1 .
- z : the maximum life cycle of system.
- T : the maximum subscribing time of user.
- B_i : public parameters, $B_i = \{D_{i,u} | D_{i,u} = a^u b^{i-u} P, \forall u \in [0, i]\}$ for $i = \{1, 2, \dots, T\}$.
- $K_{i,t}$: a class key for class C_i , $K_{i,t} = e(P, P)^{a^t b^{z-t} e_i}$, where t is a time period.
- AES: the advanced encryption standard.
- AONT: an all-or-nothing transformation.
- K_{i,t_1,t_2} : a decryption key subscribed by user, $K_{i,t_1,t_2} = e_i a^{t_1} b^{z-t_2} P$, where i denotes class C_i and t_1, t_2 denote user's subscribing time from t_1 to t_2 .

3 Proposed Scheme

In our scheme, there are three entities: data provider (DP), cloud server (CS), and user. Note that DP outsources his data to CS and it has endless storage capacity but is "honest-but-curious". In other words, it honestly follows the proposed scheme but is curious to know the content of outsourced data. The proposed scheme consists of following five phases.

Initialization. In this phase, DP provides a set of data $D = \{D_1, D_2, \dots, D_n\}$ and defines an HAC policy for D . In this policy, a set of class $C = \{C_1, C_2, \dots, C_n\}$ is defined. These classes form a directional graph $G = (V, E)$ with the relation \prec mentioned in Subsection 2.2. Then, DP chooses a bilinear map $e : G_1 \times G_1 \rightarrow G_2$ and a generator $P \in G_1$. Assume that the maximum life cycle of system is $z < q$ and the maximum subscribing time is T , where $T < z$. DP selects two random values $a, b \in Z_q^*$ and computes $B = \{B_1, B_2, \dots, B_T\}$, where $B_i = \{D_{i,u} | D_{i,u} = a^u b^{i-u} P, \forall u \in [0, i]\}$ for $i = \{1, 2, \dots, T\}$. Finally, DP publishes public parameters $\{e, G_1, G_2, q, P, B\}$.

Class Key Assignment Phase. For each class C_i , DP firstly assigns a secret value $e_i \in Z_q^*$ as a key. At each time period t , DP generates a class key $K_{i,t} = e(P, P)^{a^t b^{z-t} e_i}$. For each pair $C_j \prec C_i$, DP computes $N_{i,j,t} = AES_{K_{i,t}}(K_{j,t})$, where AES denotes the advanced encryption standard [24]. Note that $N_{i,j,t}$ is published in each time period t .

Data Outsourcing Phase. Without loss of generality, we assume that each D_i is put into class C_i for $i = 1, 2, \dots, n$. When DP wants to outsource her/his data D_i to CS in C_i , DP firstly selects a random value $k_i \in Z_q^*$ and random string r_i for D_i . Then, DP proceeds D_i with $AONT$ to get $D'_i = D_{i,1} || D_{i,2} || \dots || D_{i,\alpha'}$. Then, DP encrypts D_i to generate a ciphertext

$$\Phi_i = (\{k_i || \rho\}_{K_{i,t}}^{AES}, \{D_{i,1}\}_{k_i}^{AES} || \dots || \{D_{i,(\rho-1)}\}_{k_i}^{AES} || \{D_{i,\rho}\}_{K_{i,t}}^{AES} || \{D_{i,(\rho+1)}\}_{k_i}^{AES} || \dots || \{D_{i,\alpha'}\}_{k_i}^{AES})$$

where ρ is a random value, $1 < \rho < \alpha'$. Finally, DP sends $(ID_{D_i}, ID_{DP}, C_i, \Phi_i)$ to CS.

User Subscribing Phase. When a user U wants to subscribe class C_i from t_1 to t_2 , DP generates a decryption key $K_{i,t_1,t_2} = e_i a^{t_1} b^{z-t_2} P$ and sends it to U via a secure channel.

Decrypting Phase. Suppose that a user U subscribes class C_i in $[t_1, t_2]$. Then, she/he not only accesses D_i in C_i but also can access D_j in C_j with $C_j \prec C_i$ in any time $t \in [t_1, t_2]$. Hence, there are two cases should be concerned.

Case 1. U wants to access D_i in t . Upon receiving the request from U , CS sends the related ciphertext Φ_i to U . The user firstly compute the class key $K_{i,t} = e(K_{i,t_1,t_2}, D_{\lambda,x})$, where $t_1 + x = t = t_2 - y$, $t_2 - t_1 = \lambda = x + y$. Then, U decrypts

$\{k_i || \rho\}_{K_{i,t}}^{AES}$ to obtain k_i and ρ . Finally, all $D_{i,1}, D_{i,2}, \dots, D_{i,\alpha'}$ are obtained and thus the data D_i can be derived by using $AONT^{-1}(D_{i,1} || D_{i,2} || \dots || D_{i,\alpha'})$.

Case 2. U wants to access D_j in t . Upon receiving the request from U , CS sends the related ciphertext Φ_j to U . The user firstly computes the class key $K_{i,t}$ and then decrypts $N_{i,j,t}$ to obtain $K_{j,t}$. By the similar method in Case 1, the data D_j can be derived.

4 Security Analysis

In this section, we demonstrate the security of our scheme. Here, we consider the two types of attacks: outsider and insider attacks.

Theorem 1. *Under the security of AES and the AONT assumption, the proposed scheme is secure against outsider attacks.*

Proof. Here, there two cases should be concerned.

Case 1. A user U who does not subscribe any class C_i from DP cannot compute the class key $K_{i,t}$ because U has no the decryption key K_{i,t_1,t_2} . Furthermore, U cannot generates a fake decryption key K'_{i,t_1,t_2} in some time interval $[t_1, t_2]$ because e_i, a, b are secret values kept by DP. In other aspect, U cannot break the ciphertext Φ_i directly to obtain D_i under the security of AES and the AONT assumption.

Case 2. CS cannot obtain D_i from the ciphertext Φ_i . This case is a special case of Case 1.

Theorem 2. *Under the security of AES and the AONT assumption, the proposed scheme is secure against insider attacks.*

Proof. Firstly, we consider a simple case that a user U who subscribe class C_i in $[t_1, t_2]$ cannot access D_i in time t_3 for $t_3 > t_2$ or $t_3 < t_1$. Here, U has got the key $K_{i,t_1,t_2} = e_i a^{t_1} b^{z-t_2} P$. In order to access D_i in time t_3 , U must obtain the class key $K_{i,t_3} = e(P, P)^{a^{t_3} b^{z-t_3} e_i}$. Hence, U may find a point $D = a^{t_3-t_1} b^{t_2-t_3} P \in G_1$ such that $K_{i,t_3} = e(K_{i,t_1,t_2}, D)$. However, it is impossible because a and b are secret values kept by DP.

Then, we consider the colluding attacks. Assume that there exist two users U_1 and U_2 who collude to access the data which is not subscribed by them. Here, there four cases should be concerned.

Case 1. Assume that U_1 subscribes C_j from t_1 to t_2 and U_2 subscribes C_k in the same time interval, where $C_j \prec C_i$ and $C_k \prec C_i$. They want to compute the class key $K_{i,t}$ which can access D_i in the class C_i from t_1 to t_2 . Now, U_1 has the key $K_{j,t_1,t_2} = e_j a^{t_1} b^{z-t_2} P$ and U_2 has the key $K_{k,t_1,t_2} = e_k a^{t_1} b^{z-t_2} P$. However, it is impossible to compute $K_{i,t_1,t_2} = e_i a^{t_1} b^{z-t_2} P$ because e_i, a, b are secret values kept by DP.

Case 2. Assume that U_1 subscribes C_j from t_1 to t_3 and U_2 subscribes C_k from t_2 to t_4 , where $C_j \prec C_i$, $C_k \prec C_i$, and $t_1 < t_2 < t_3 < t_4$. They want to compute the class key $K_{i,t}$ which can access D_i in the class C_i from t_2 to t_3 . Now, U_1

has the key $K_{j,t_1,t_3} = e_j a^{t_1} b^{z-t_3} P$ and U_2 has the key $K_{k,t_2,t_4} = e_k a^{t_2} b^{z-t_4} P$. However, it is impossible to compute $K_{i,t_2,t_3} = e_i a^{t_2} b^{z-t_3} P$ because e_i, a, b are secret values kept by DP.

Case 3. From Case 1, assume that the cloud server (CS), U_1 , and U_2 collude. They want to access D_i in the class C_i from t_1 to t_2 . By Case 1, to compute $K_{i,t_1,t_2} = e_i a^{t_1} b^{z-t_2} P$ is infeasible. By Theorem 1, they cannot break the ciphertext Φ_i directly to obtain D_i under the security of *AES* and the *AONT* assumption.

Case 4. From Case 2, assume that the cloud server (CS), U_1 , and U_2 collude. They want to access D_i in the class C_i from t_2 to t_3 . By Case 2, to compute $K_{i,t_2,t_3} = e_i a^{t_2} b^{z-t_3} P$ is infeasible. By Case 3, they cannot break the ciphertext Φ_i directly to obtain D_i under the security of *AES* and the *AONT* assumption.

5 Performance Analysis

For convenience to evaluate the performance of our scheme, we first define the following notations:

- TG_e : The time of executing a bilinear pairing operation, $e : G_1 \times G_1 \rightarrow G_2$.
- TG_{mul} : The time of executing a scalar multiplication operation of point in G_1 .
- T_{exp} : The time of executing a modular exponentiation operation.
- T_{AES} : The time of executing the *AES* algorithm.
- l : The number of blocks for a outsourced file.
- d : The path length between the subscribing class and its lower level classes.

Here, we demonstrate the executing time in each phase of our scheme in Table 1.

Table 1. The executing time in each phase of our proposed scheme

Phases	Executing time
Data outsourcing	$n(l+1)T_{AES}$
User subscribing	$TG_{mul} + 2T_{exp}$
Decrypting for subscribing class	$TG_e + (l+1)T_{AES}$
Decrypting for lower class of subscribing class	$TG_e + (l+d)T_{AES}$

6 Conclusions

In this paper, we have proposed the first time-bound hierarchical key management scheme in cloud computing. Our scheme does not require any tamper-resistant devices and is suitable for cloud environments. The security analysis is demonstrated that our scheme is secure against outsider and insider attacks. For the future work, we will design a new scheme for discrete time period.

Acknowledgments. This work is supported by Shenzhen Peacock Project of China (No. KQC201109020055A), Shenzhen Strategic Emerging Industries Program of China (No. ZDSY20120613125016389 and No. JCYJ20120613151032592), and National Natural Science Foundation of China (No. 61100192).

References

1. Tang, Y., Lee, P., Lui, J., Perlman, R.: Secure overlay cloud storage with access control and assured deletion. *IEEE Transactions on Dependable and Secure Computing* 9(6), 903–916 (2012)
2. Jung, T., Li, X.Y., Wan, Z., Wan, M.: Privacy preserving cloud data access with multi-authorities. In: *IEEE INFOCOM*, pp. 2625–2633. IEEE Press, New York (2013)
3. Chu, C.K., Chow, S.S.M., Tzeng, W.G., Zhou, J., Deng, R.H.: Key-aggregate cryptosystem for scalable data sharing in cloud storage. *IEEE Transactions on Parallel and Distributed Systems* 25(2), 468–477 (2014)
4. Akl, S.G., Taylor, P.D.: Cryptographic solution to a problem of access control in a hierarchy. *ACM Transactions on Computer Systems* 1(3), 239–248 (1983)
5. Jiang, T., Zheng, S., Liu, B.: Key distribution based on hierarchical access control for conditional access system in DTV broadcast. *IEEE Transactions on Consumer Electronics* 50(1), 225–230 (2004)
6. Atallah, M.J., Blanton, M., Fazio, N., Frikken, K.B.: Dynamic and efficient key management for access hierarchies. In: *12th ACM Conference on Computer and Communications Security*, pp. 190–201. ACM Press, New York (2005)
7. Kayem, A.V.D.M., Martin, P., Akl, S.G.: Heuristics for improving cryptographic key assignment in a hierarchy. In: *21st International Conference on Advanced Information Networking and Applications Workshops*, pp. 531–536. IEEE Press, New York (2007)
8. Tzeng, W.G.: A time-bound cryptographic key assignment scheme for access control in hierarchy. *IEEE Transactions on Knowledge and Data Engineering* 14(1), 182–188 (2002)
9. Yi, X., Ye, Y.: Security of Tzeng’s time-bound key assignment scheme access control in a hierarchy. *IEEE Transactions on Knowledge and Data Engineering* 15(4), 1054–1055 (2003)
10. Chien, H.Y.: Efficient time-bound hierarchical key assignment scheme. *IEEE Transactions on Knowledge and Data Engineering* 16(10), 1301–1304 (2004)
11. Yi, X.: Security of Chien’s efficient time-bound hierarchical key assignment scheme. *IEEE Transactions on Knowledge and Data Engineering* 17(9), 1298–1299 (2005)
12. Yeh, J.H.: An RSA-based time-bound hierarchical key assignment scheme for electronic article subscription. In: *14th ACM International Conference on Information and Knowledge Management*, pp. 285–286. ACM Press, New York (2005)
13. Wang, S.Y., Lih, C.S.: Merging: an efficient solution for a time-bound hierarchical key assignment scheme. *IEEE Transactions on Dependable and Secure Computing* 3(1), 91–100 (2006)
14. Ateniese, G., Santis, A.D., Ferrara, A.L., Masucci, B.: Provably-secure time-bound hierarchical key assignment schemes. In: *13th ACM Conference on Computer and Communications Security*, pp. 288–297. ACM Press, New York (2006)

15. Sui, Y., Maino, F., Guo, Y., Wang, K., Zou, X.: An efficient time-bound access control scheme for dynamic access hierarchy. In: 5th International Conference on Mobile Ad-hoc and Sensor Networks, pp. 279–286. IEEE Press, New York (2009)
16. Chen, C.M., Wu, T.Y., He, B.Z., Sun, H.M.: An efficient time-bound hierarchical key management scheme without tamper-resistant devices. In: 1st International Conference on Computing, Measurement, Control and Sensor Network, pp. 285–288. IEEE Press, New York (2012)
17. Tseng, Y.M., Yu, C.H., Wu, T.Y.: Towards scalable key management for secure multicast communication. *Information Technology and Control* 41(2), 173–182 (2012)
18. Chen, Y.-R., Chu, C.-K., Tzeng, W.-G., Zhou, J.: CloudHKA: a cryptography approach for hierarchical access control in cloud computing. In: Jacobson, M., Locasto, M., Mohassel, P., Safavi-Naini, R. (eds.) ACNS 2013. LNCS, vol. 7954, pp. 37–52. Springer, Heidelberg (2013)
19. Boneh, D., Franklin, M.: Identity-based encryption from the Weil pairing. In: Kilian, J. (ed.) CRYPTO 2001. LNCS, vol. 2139, pp. 213–229. Springer, Heidelberg (2001)
20. Chen, L., Cheng, Z., Smart, N.P.: Identity-based key agreement protocols from pairings. *International Journal of Information Security* 6(4), 213–241 (2007)
21. Wu, T.Y., Tseng, Y.M.: An ID-based mutual authentication and key exchange protocol for low-power mobile devices. *The Computer Journal* 53(7), 1062–1070 (2010)
22. Sandhu, R.S., Samarati, P.: Access control: principle and practice. *IEEE Communications Magazine* 32(9), 40–48 (1994)
23. Rivest, R.L.: All-or-nothing encryption and the package transform. In: Biham, E. (ed.) FSE 1997. LNCS, vol. 1267, pp. 210–218. Springer, Heidelberg (1997)
24. Advanced Encryption Standard (AES),
<http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf>

Shape Estimation from 3D Point Clouds

Jingyong Su¹ and Lin-Lin Tang²

¹ Texas Tech University

jingyong.su@ttu.edu

² Harbin Institute of Technology Shenzhen Graduate School

linlintang2009@gmail.com

Abstract. The problem of estimating a shape in a 3D point cloud data is important due to its general applicability in image analysis, computer vision, and graphics. It is challenging because the data is typically noisy, cluttered, partly missing and unordered. We address shape estimation using a template object under a fully statistical model, where the data is assumed to be modeled using a Poisson process on the object's boundary (surfaces), corrupted by additive noise and a clutter process. Using analytical likelihood function dictated by the model, we optimize over pose and scale associated with hypothesized templates and estimate most likely shapes in observed point clouds under given shape hypotheses. We demonstrate this framework using examples of 2D and 3D shape estimation in simulated and real data.

1 Introduction

The 3D reconstruction from point clouds is a topic of major interest in computer vision. The emergence of laser/lidar sensors, reliable multi-view stereo techniques and more recently consumer depth cameras have brought point clouds to the forefront as a data format useful for a number of applications in areas such as reverse engineering, product design, medical appliance design and archeology, among others. Since point clouds are so general, as they contain no interpretation of the data, they are broadly applicable in different scientific domains. Furthermore, the acquisition and processing of digital 3D point clouds has received increasing attention over the last few years. While visualization of very detailed and complex point clouds has become possible, interaction capabilities on a semantic level are still very limited. Even tasks as basic as selecting all windows in a scan of a house currently require a disproportional amount of user interaction. This is due to the fact that the acquired raw data does not provide any structure let alone semantic information. Therefore the extraction of structured shapes from 3D point clouds is an important topic for a wide field of applications.

The problem of shape estimation in point clouds without additional information is nearly impossible. In this paper, we restrict to a subproblem of finding a shape when a template of its shape class is given to us. The actual shape may differ from the template due to articulation within class variation and missing parts.

What makes this subproblem difficult? Here are some of the issues: (1) **Unknown Pose and Scale:** A shape can be present in arbitrary pose and scale as shown in Fig. 1 and one does not know these variables a-priori. The key is to find global solutions for unknown variables, especially the pose and scale. (2) **Noise and Clutter:** There is invariably some observation noise associated with shape measurements and also presence of points that belong to either the background or other objects, termed *clutter* in this paper, as shown in Fig. 1. (3) **Missing Parts:** In the presence of occlusion, one or more parts of an object may be absent from the data. In this setting it is natural to develop statistical models, and seek efficient and global solutions for estimating unknown variables. Note that this problem is different from the problem of comparing unlabeled point patterns which has been addressed by a number of landmark papers, including [10,7]. In our case, amongst the two objects being compared, one is a well-defined shape while another is a point cloud, while in these papers both the objects can be point clouds. Also, while most previous works focus on registration between two point clouds, our problem is to actually estimate a shape.

1.1 Past Work

There has been much work on object recognition from 2D and 3D point clouds. They can be divided into a few broad categories.

The first category of papers detects shapes from 2D point clouds. [15] presents a statistical approach for the identification of objects in digital images. Using Procrustes analysis, a point distribution model is fitted on a set of training images and used as a prior distribution for the shape of a deformable template. A recent paper by [16] develops a Bayesian approach for shape classification in 2D point clouds. Here the authors estimate the posterior probability of a given shape by integrating over unknown variables such as pose, scale, and point labels using Monte Carlo method. [17] presents a likelihood-based framework for shape detection in 2D point clouds. A generalized likelihood ratio test is developed using Poisson model-based approach.

The second category of papers works on 3D points clouds and surface reconstruction including [9,1,11,13,6,12]. Most methods call for some kind of connectivity information and are not well equipped to deal with a large amount of outliers. [12] present a method of modeling cities form point clouds. 3D-primitives such as planes, cylinders, spheres or cones are combined with mesh-patches. A region growing approach has also been used to detect planes in 3D point clouds by [18]. This approach often delivers a superior segmentation but still suffers from the problems on noisy data. In computer graphics, [5] have recently proposed a general variational framework for approximation of surfaces by planes, which was extended to a set of more elaborate shape proxies by [19].

The iterative closest point (ICP) algorithm by [2] is a simple method that uses the nearest-neighbor relationship to assign a binary correspondence at each step. This estimate of the correspondence is then used to refine the transformation, and vice versa. [8] finds the locations of target objects using single spin image matching and then retrieves the orientation and quality of the match using the

ICP algorithm. [4] proposes a point matching algorithm for non-rigid registration. They develop an algorithm with the thin-plate splines as the parameterization of the non-rigid spatial mapping and the softassign for the correspondence. [3] presents a technique that uses a vector space representation of shape (3D Morphable Model) to infer missing vertex coordinates.

1.2 Our Approach

We use a fully statistical framework for estimating *pre-determined shapes* in point clouds introduced in [17]. It is a model-based approach where the data is modeled using a Poisson process on the object’s boundary, corrupted by an additive noise and a clutter process. Using analytical likelihood functions dictated by the model, we optimize pose and scale associated with hypothesized objects and estimate most likely shapes in observed point clouds under given shape hypotheses. To guarantee the global optima, we use grid search on a subset of variables despite additional computational cost.

2 Problem Formulation

We consider the following problem: We are given a point cloud $\mathbf{y} = \{\mathbf{y}_i \in \mathbb{R}^3, i = 1, 2, \dots, m\}$ in a domain U and we want to develop a statistical framework for estimating a pre-determined shape contained in this set. By a shape we mean a parametrized surface but at an arbitrary pose, scale, and parameterization. Shape is a characteristic that is invariant to similarity transformations, but when a shape occurs in a scene, it has a specific scale, position, and orientation. From the perspective of shape estimation, these variables have to be estimated. Also, we take a model-based approach.

To illustrate the problem, some examples of 3D point clouds are shown in Fig. 1. We are interested in finding and estimating the shape of a “fish” present in these clouds. A quick inspection ascertains the presence and pose of a fish in the first two cases, even though the second one is more cluttered than the first one, but the situation is not so clear for the last case. We would like to perform this estimation automatically.

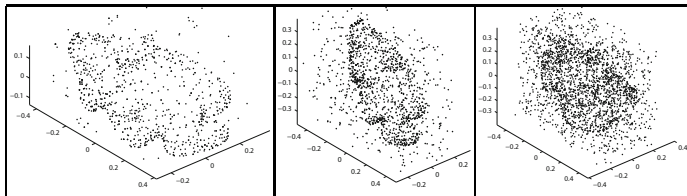


Fig. 1. Examples of 3D cluttered point clouds

2.1 Likelihood Estimation

The points present in a given point cloud can be one of two types: (i) points belonging to a shape and (ii) points associated with the background clutter. We will propose an observation model for each of them separately.

In order to better explain the model description, we will start with a simpler problem where we seek a specific object, i.e. fixed shape, pose, and scale. Let $\beta : D \rightarrow \mathbb{R}^3$ be a parameterized object, where D is a domain for the parameterization. We make the following modeling choices:

1. **Points belonging to β :** We assume that these points are realizations of a Poisson process on the parameterized object β . Let $\gamma : D \rightarrow \mathbb{R}_{\geq 0}$ be the intensity function of the Poisson process along β ; the number of points generated from any part of the object is a Poisson random variable with mean being the integral of γ on that part. In particular, k , the total number of points belonging to the object, is a Poisson random variable with mean $\Gamma = \int_D \gamma(s) ds$. Let the points sampled from β be denoted by $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k]$, $\mathbf{x}_i \in \mathbb{R}^3$. The actual observations \mathbf{y}_i are assumed to be noisy versions of \mathbf{x}_i . For a given point \mathbf{x}_i , the locations of \mathbf{y}_i are independent of each other with the identical density $f(\mathbf{y}_i|\mathbf{x}_i)$.

2. **Points associated with clutter:** This subset of observations, independent of the first subset, comes from the clutter and we model them as realizations of a Poisson process with the intensity $\lambda : U \subset \mathbb{R}^3 \rightarrow \mathbb{R}_{\geq 0}$, where U is the region containing observed points, e.g. $U = [a, b]^3$ for 3D point clouds. Let $\Lambda = \int_U \lambda(y) dy$. The full observation \mathbf{y} can now be modeled as a Poisson process with the intensity function: $\xi(y) = \int_D f(y|\beta(s))\gamma(s) ds + \lambda(y)$. The probability density function of \mathbf{y} , given β, γ, λ for a fixed m , is given by: $P_m(\mathbf{y}|\beta, \gamma, \lambda) = (\prod_{i=1}^m \xi(\mathbf{y}_i))e^{-\Lambda - \Gamma}$, where m is the total number of points in the data.

So far the unknown parameters are full functions and that involves tremendous computational complexity. We will simplify the evaluation of $P_m(\mathbf{y}|\beta, \gamma, \lambda)$ by making some additional assumptions as follows:

1. The noise added to the points sampled from β is i.i.d. Gaussian with mean zero and variance $\sigma^2 I_{3 \times 3}$. Therefore, the conditional density $f(y|x)$ takes the form $\frac{1}{(2\pi)^{3/2}\sigma^3} e^{-\frac{1}{2\sigma^2}\|y-x\|^2}$ for $y, x \in \mathbb{R}^3$.
2. Both the Poisson intensities are constant, i.e., $\lambda(y) = \lambda$ and $\gamma(s) = \gamma$. In this case $\Lambda = \lambda \int_U dy$ and $\Gamma = \gamma \int_D ds$. To simplify the discussion, we scale both the integrals to be one such that $\Lambda = \lambda$ and $\Gamma = \gamma$.

With these assumptions, the probability density function simplifies to: $P_m(\mathbf{y}|\beta, \gamma, \lambda) = e^{-\gamma - \lambda} (\prod_{i=1}^m (\lambda + \gamma \alpha_\sigma(\mathbf{y}_i)))$. $\alpha_\sigma : \mathbb{R}^3 \rightarrow \mathbb{R}_+$ is a scalar map given by $\alpha_\sigma(\mathbf{y}_i) = \frac{1}{(2\pi)^{3/2}\sigma^3} \int_D e^{-\frac{1}{2\sigma^2}\|\mathbf{y}_i - \beta(s)\|^2} ds$. Notice that $\alpha_\sigma(\mathbf{y}_i)$ is high if a point \mathbf{y}_i is close to the object β , with the closeness being measured relative to the scale σ .

Let $\theta = [\gamma, \lambda, \sigma] \in \mathbb{R}^3$ denote three unknown parameters associated with the shape. Then, define a log-likelihood function $H : \mathbb{R}^3 \rightarrow \mathbb{R}_+$ given by $H(\theta) = \log(P_m(\mathbf{y}|\beta, \gamma, \lambda)) = -\gamma - \lambda + \sum_{i=1}^m \log(\lambda + \gamma \alpha_\sigma(\mathbf{y}_i))$, and $\hat{\theta} = \operatorname{argmax}_\theta H(\theta)$ is the MLE of the nuisance parameters.

2.2 MLE of θ

Given a point cloud \mathbf{y} , and an object β , we need to solve for the MLE $\hat{\theta}$ and we do that using a gradient approach. The parameter θ has three components. Of these, we search exhaustively for the parameter σ and use a gradient-based approach to search over the remaining two γ and λ . For each value of σ in a certain range, say $[\sigma_l, \sigma_u]$, we maximize H over the pair (γ, λ) . The log-likelihood function H is a concave function in λ and γ (for a fixed σ). The gradients of estimating parameters λ and γ are given by:

$$\frac{\partial H}{\partial \lambda} = -1 + \sum_{i=1}^m \left(\frac{1}{\lambda + \gamma \alpha_\sigma(\mathbf{y}_i)} \right), \quad \frac{\partial H}{\partial \gamma} = -1 + \sum_{i=1}^m \left(\frac{\alpha_\sigma(\mathbf{y}_i)}{\lambda + \gamma \alpha_\sigma(\mathbf{y}_i)} \right). \quad (1)$$

The algorithm is summarized below:

Algorithm 1 (MLE of θ):

- For each $\sigma \in [\sigma_l, \sigma_u]$ perform the following:
 1. Set $k = 0$ and initialize the pair $[\gamma_k, \lambda_k]$ with random values in the range $[0, m]$.
 2. Update the estimates using: $\begin{bmatrix} \gamma_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} \gamma_k \\ \lambda_k \end{bmatrix} + \delta \begin{bmatrix} \frac{\partial H}{\partial \gamma}(\gamma_k, \lambda_k) \\ \frac{\partial H}{\partial \lambda}(\gamma_k, \lambda_k) \end{bmatrix}$, for a small $\delta > 0$.
 3. If the norm of the gradient vector is small, then stop. Else, set $k = k + 1$ and return to step 2.
- Set the current values to be $(\hat{\gamma}(\sigma), \hat{\lambda}(\sigma))$.

Define the MLE $\hat{\theta}$ to be $(\hat{\gamma}(\hat{\sigma}), \hat{\lambda}(\hat{\sigma}), \hat{\sigma})$ where $\hat{\sigma} = \operatorname{argmax}_{\sigma \in [\sigma_l, \sigma_u]} H(\hat{\gamma}(\sigma), \hat{\lambda}(\sigma), \sigma)$.

2.3 MLE of Pose and Scale Variables

So far we have assumed a fixed surface β , but of course a surface can be present at an arbitrary pose and scale. Therefore, β can have variable position, rotation, and scale. Let $O \in SO(3)$ denote the orientation, $T \in \mathbb{R}^3$ denote its translation, and $\rho \in \mathbb{R}_+$ denote its scale. First let β_0 be a standardized surface (center of mass is zero, and major axes are aligned with canonical axes) with a fixed shape and define $\beta = \rho O \beta_0 + T$ be a transformed version of that surface. When we allow unknown transformations, denoted by $\omega = (\rho, O, T)$, the cost function H keeps the same form, except the function α_σ now depends on these transformation variables, i.e., $\alpha_\sigma(\mathbf{y}_i | \omega) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^3 \int_{S^2} e^{-\frac{1}{2\sigma^2} \|\mathbf{y}_i - \rho O \beta_0(s) - T\|^2} ds$. These additional variables are also estimated, along with θ , using maximum-likelihood estimation.

The function H may have local maximums in ω space and we try to improve performance using several initial values in the scale variable. The gradient algorithm is summarized below:

Algorithm 2 (MLE of ω):

- Initialize \mathbf{y} : Translate the center of \mathbf{y} to the origin and apply singular value decomposition (SVD) on it to initialize the rotation, i.e., initialize translation vector to be zero and rotation matrix to be identity.
- For the scale variable ρ , choose a number of values in spaces uniformly in an interval, say $[\rho_l, \rho_u]$.
- Gradient-based search: For each scale value, apply the gradient method. Finally, select the solution that results in the largest H value.

2.4 Shape Estimation

Now, given a 3D point cloud \mathbf{y} and the matched shape surface, i.e. $\beta = (\beta_0$ at the estimated pose and scale), we first choose points whose value of $\alpha_{\hat{\sigma}}$ is larger than a threshold. The threshold is also determined by trial and error. Then using Hungarian algorithm, for each of chosen points, we find the unique nearest neighbor on the surface β . Now the remaining problem is equivalent to reconstructing a triangulated surface only using these selected neighbors in β . We use the algorithm of vertex decimation, an iterative simplification algorithm originally proposed in [14] to remove unwanted points in β . By doing this, the points on β which are not the nearest neighbors of points selected in \mathbf{y} are removed and we reconstruct a triangulated surface only using the neighbors. Finally, we borrow the connection information from it and estimate a shape from the selected points.

3 Experimental Results

In this section, we will show some examples of full shape estimation and partial shape estimation separately.

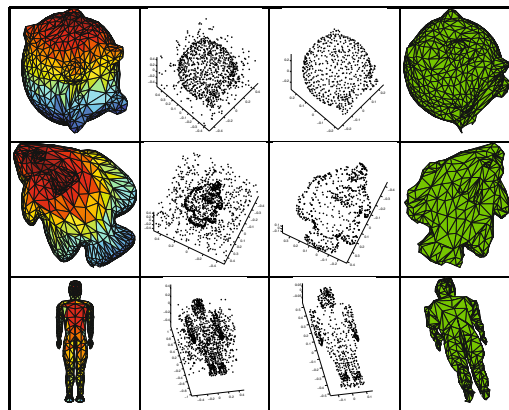


Fig. 2. Examples of full shape estimation

3.1 Full Shape Estimation

Given a shape template, we simulate a point cloud by sampling points from the arbitrarily rotated and scaled template with additional background clutter. Then we select points using the framework and estimate a shape from them, shown in Fig. 2. Each row lists shape template, simulated point cloud, selected points and estimated shape.

3.2 Partial Shape Estimation

Given a point cloud, which is simulated from a part of template, we can still optimize over pose and scale variables using our framework and estimate a part of shape from the cloud. Fig. 3 illustrates that the method can also estimate shape that are only partially present in the data. Each column lists simulated point cloud from part of the shape template, estimated template in the cloud and estimated shape from the cloud.

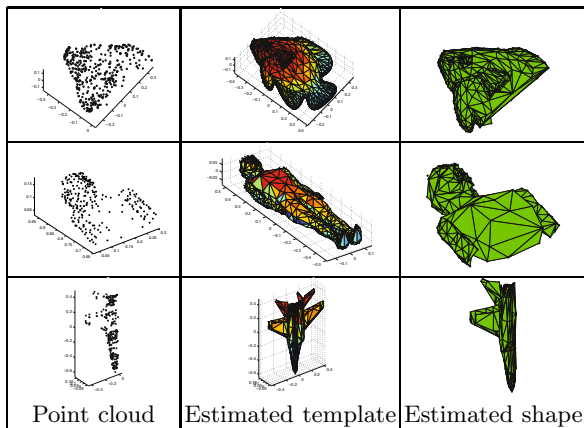


Fig. 3. Examples of partial shape estimation

4 Conclusion

We have presented a fully statistical framework for estimating shapes in 3D cluttered point cloud and have demonstrated it using multiple examples. This model is based on a composite Poisson process: one for points generated from the shape and another for points belonging to the background clutter. This model allows computation of a log-likelihood function and optimizing over pose and scale variables.

References

1. Alexa, M., Behr, J., Cohen-Or, D., Fleishman, S., Levin, D., Silva, C.T.: Point set surfaces. In: Proc. Visualization 2001, pp. 21–28 (2001)
2. Besl, P., McKay, H.: A method for registration of 3-D shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 14(2), 239–256 (1992)
3. Blanz, V., Mehl, A., Vetter, T., Seidel, H.: A statistical method for robust 3D surface reconstruction from sparse data. In: Proc. the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission, pp. 293–300 (2004)
4. Chui, H., Rangarajan, A.: A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding* 89(2-3), 114 – 141 (2003)
5. Cohen-Seiner, D., Alliez, P., Desbrun, M.: Variational shape approximation. *ACM Trans. on Graphics* 23, 905–914 (2004)
6. Dey, T.K., Goswami, S.: Provable surface reconstruction from noisy samples. *Journal of Computational Geometry: Theory and Applications* 35, 124–141 (2006)
7. Dryden, I.L., Hirst, J.D., Melville, J.L.: Statistical analysis of unlabeled point sets: Comparing molecules in chemoinformatics. *Biometrics* 63(1), 237–251 (2007)
8. Halma, A., ter Haar, F., Bovenkamp, E., Eendebak, P., van Eekeren, A.: Single spin image-ICP matching for efficient 3D object recognition. In: Proc. ACM workshop on 3D object retrieval, pp. 21–26 (2010)
9. Hoppe, H., DeRose, T., Duchamp, T., McDonald, J., Stuetzle, W.: Surface reconstruction from unorganized points. In: Proc. ACM SIGGRAPH, pp. 71–78 (1992)
10. Kent, J., Mardia, K., Taylor, C.: Matching problems for unlabeled configurations. In: Aykroyd, R.G., Barber, S., Mardia, K.V. (eds.) *LASR 2004 Proc. Bioinformatics, Images, and Wavelets*, pp. 33–40 (2004)
11. olluri, R., Shewchuk, J.R., O’Brien, J.F.: Spectral surface reconstruction from noisy point clouds. In: Proc. Geometry Processing (Eurographics/ ACM SIGGRAPH), pp. 11–21 (2004)
12. Lafarge, F., Mallet, C.: Building large urban environments from unstructured point data. In: *ICCV*, pp. 1068 –1075 (2011)
13. Mederos, B., Amenta, N., Velho, L., de Figueiredo, L.H.: Surface reconstruction from noisy point clouds. In: Proc. Geometry Processing (Eurographics/ ACM SIGGRAPH), pp. 53–62 (2005)
14. Schroeder, W.J.: A topology modifying progressive decimation algorithm. In: Proc. Visualization 1997, pp. 205–212 (1997)
15. de Souza, K.M.A., Kent, J.T., Mardia, K.V.: Stochastic templates for aquaculture images and a parallel pattern detector. *Journal of the Royal Statistical Society Series C* 48(2), 211–227 (1999)
16. Srivastava, A., Jermyn, I.H.: Looking for shapes in cluttered, two-dimensional point clouds. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 31(9), 1616–1629 (September, 2009)
17. Su, J., Zhu, Z., Srivastava, A., Huffer, F.: Detection of shapes in 2d point clouds generated from images. In: *ICPR*, pp. 2640–2643 (2010)
18. Verma, V., Kumar, R., Hsu, S.: 3D building detection and modeling from aerial lidar data. In: *CVPR*, vol. 2, pp. 2213 – 2220 (2006)
19. Wu, J., Kobbelt, L.: Structure recovery via hybrid variational surface approximation. *Computer Graphics Forum* 24, 277–284 (2005)

Deterministic Data Sampling Based on Neighborhood Analysis

Sarka Zehnalova, Milos Kudelka, and Jan Platos

VSB - Technical University of Ostrava, Czech Republic,
17. listopadu 15, 708 33, Czech Republic
{sarka.zehnalova.st,milos.kudelka,jan.platos}@vsb.cz

Abstract. The amount of large-scale real data around us is increasing in size very quickly, as is the necessity to reduce its size by obtaining a representative sample. Such sample allows us to use a great variety of analytical methods, the direct application of which on original data would be unfeasible. There are many methods used for different purposes and with different results. In this paper, we outline a simple, flexible and straightforward approach based on analyzing the nearest neighbors that is generally applicable. This feature is illustrated in experiments with synthetic and real-world datasets. The properties of the representative sample show that the presented approach maintains very well internal data structures (e.g. clusters and density). The key technical parameters of the approach are low complexity and high scalability.

Keywords: sampling, data mining, density bias, nearest neighbor.

1 Introduction

In the area of big data, sampling may help facilitate knowledge discovery from large-scale datasets. Using analytical methods it is possible to find patterns and regularities in a sample that is significantly smaller than the original dataset. To be able to validate the observed patterns in the original data, it is necessary to have a representative sample which retains certain statistical properties.

In the field of large datasets with vector data, it is usually the assessment of the extent to which the sample maintains clusters and their density [5,8]. Generally, the approaches can be divided into two groups; unbiased (uniform random sampling) and biased. The approach described in this paper is one of the biased methods and is based on the selection of representatives, i.e. objects that in their surroundings play a more important role than others.

Our method has three important aspects that will be subsequently described in the paper. The first is determinism. A representative sample is uniquely determined by the parameters specified in the method. The second aspect is locality. The method works with the neighborhood of an object (in the meaning of a small distance between objects). That implies the scalability of the presented algorithm, which is the third aspect.

We illustrate the usage of our method in experiments with small 2-dimensional datasets, and also with a large real-world dataset. We provide a visual comparison of our sampling method with uniform random sampling.

The paper is organized as follows: In section 2, we discuss the related work. The proposed method is presented in Section 3. In section 4, we focus on the experiment and on its results. Section 5 concludes the paper.

2 Sampling Methods

Sampling [1,4] has already been applied in many areas to various types of data. The goal is to reduce the original dataset to a more manageable size. For large n-dimensional dataset sampling methods have been developed in order to optimize data mining tasks [9,13], such as clustering or outlier detection. The approaches can be divided into two groups; unbiased and biased. Each is suitable in different applications.

2.1 Uniform Random Sampling

In uniform random sampling, every data point has the same probability of being selected for the sample. This approach has been used extensively in database and data mining tasks [2,6,11]. Techniques have been developed to collect a sample in one single sequential pass through the file [10]. In the BIRCH [12] clustering algorithm, a random sample is selected as the initial step.

2.2 Biased Sampling

In biased sampling, every data point has a different probability of being selected for the sample. This approach is suitable for clustering tasks where the dataset includes clusters of different sizes. With uniform random sampling, small clusters are likely to be omitted. In this case, density-biased sampling [7,8] gives better results since sparse groups (areas) are given a higher probability of being included in the sample. Kollios et al. [5] also used density-biased sampling to speed up cluster and outlier detection. They use kernel density estimation to obtain local density which is calculated for every point. Kerdprasop et al. [3] proposed a modification of the k-means clustering algorithm where density-biased reservoir sampling is used.

3 Deterministic Sampling Based on Neighborhood

In this section, we define the basic ideas behind the proposed algorithm. All of the above-mentioned methods produce a random sample from an original dataset, while our approach is deterministic. Therefore, the resulting representative sample is determined by the configuration of the algorithm.

3.1 Sampling Algorithm

Our approach to obtaining a representative sample of a dataset is inspired by the method of finding nearest neighbors. The algorithm is based on the idea that objects which are the nearest neighbors of other objects are the important ones in a dataset. It is a local-oriented algorithm, which reduces the given dataset to its sample. We use a very simple function for defining the representatives of the dataset. This function is based on the neighborhood analysis and depends on the distance between objects.

In this paper we use vector datasets, so the object is an n -dimensional vector. We use the standard Euclidean distance as a distance measure. There exists the function $N(x)$ (Neighborhoods) which for any object x returns the number of objects to which x is close (i.e. number of neighborhoods the object x belongs to). The size of the neighborhood is defined by the maximum distance from the examined object. In this paper, the neighborhood is an n -dimensional sphere with defined radius. There exists the function $NN(x)$ (Nearest Neighbor) which for any object x returns the number of objects for which x is the closest object in their neighborhood.

The “is representative” function $R(x)$ is defined as a Boolean function based on the ratio of the *Nearest Neighbor* and the logarithm of the *Neighborhoods*. The base of the logarithm b is a parameter of the function $R(x)$. Formally, we may write this function as follows:

$$R(x) = \left(\frac{NN(x)}{\log_b(N(x))} \geq 1 \right)$$

Algorithm 1 describes obtaining the representative sample of a dataset.

3.2 Algorithm Complexity

The complexity of the algorithm may be clearly extracted from the pseudo-code in Algorithm 1. If we suppose that the dataset D contains N objects and the average size of the neighborhood is M , then the complexity of the algorithm is $O(NM)$. Usually, we may assume that $M \ll N$ so the complexity is linear. This is done because of the locality of the algorithm. If we think more deeply about the algorithm, we see that the complexity is highly affected by the complexity of the neighborhood discovery. This is affected by the distance and proximity functions, but if we suppose that these two functions follow the locality, in a dataset with vector data we then know only the information about each vector itself, but we have no information about its neighbors. So we must use a certain data structure which enables fast neighborhood exploration. Many such structures have been developed in the past, such as R-Tree and KD-Tree and their variants, or when the data have a small dimension we may use Quadrant tree. These structures allow us to find neighbors in constant or, in the worst case, logarithmic time, so the efficiency of the algorithm is still very good.

```

input : dataset  $D$ , radius  $r$ , logarithm base  $b$ 
output: sample  $S$ 
foreach object  $o \in D$  do
  | initialize the counters  $N(o)$  and  $NN(o)$  to zero
end
foreach object  $o \in D$  do
  | find neighborhood  $\varepsilon(o)$  according to the radius  $r$ 
  | foreach object  $x \in \varepsilon(o)$  do
  | | increase the counter  $N(x)$ 
  | end
  | find set  $nn(o) \subset \varepsilon(o)$  of the nearest objects to  $o$ 
  | foreach object  $y \in nn(o)$  do
  | | increase the counter  $NN(y)$ 
  | end
end
foreach object  $o \in D$  do
  | if  $R(o)$  ( $o$  is representative) then
  | | add  $o$  to  $S$ 
  | endif
end

```

Algorithm 1. Sampling algorithm

4 Experimental Evaluation

The experiments with vector data focused mainly on the ability of the proposed algorithm to reduce data while preserving important features such as clusters and local density.

When we deal with vector data and metric spaces in the way we defined our distance function, we have two parameters which may be tuned according to the expected result - logarithm base and radius. These two parameters are slightly related but each work in a different way, as the experiments shows. We tested many settings of these parameters and we present the results achieved with different settings of the algorithm.

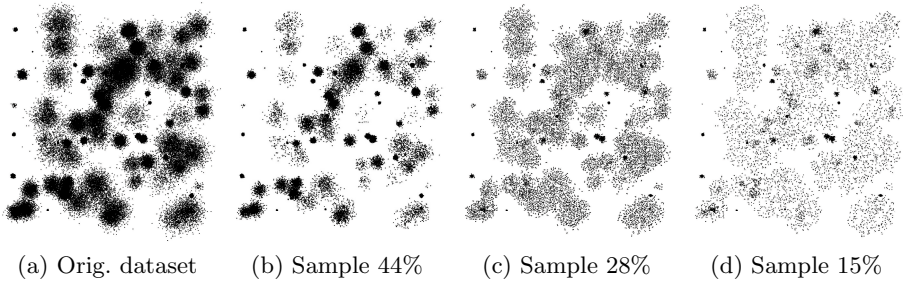
4.1 Experiment 1: Example Datasets

The first (synthetic) dataset contains several more or less dense, random sized, clusters in random locations. This dataset has 100,000 objects. Additionally to the distance function as a Euclidean distance, we discretize the distance with step 100. We experimentally set the base of the logarithm to 4 and the radius of the neighborhood to 50, 100, and 200 units. A summary of the experiment is depicted in Table 1 and a visualization of the data is shown in Figures 1a-1d.

As may be seen from the figures, the first sample (see Figure 1b) preserves the cluster centers precisely; especially the densest areas are preserved. The objects on the border of clusters and objects with a greater distance from the cluster centers are removed. The second sample (see Figure 1c) with greater radius shows a different aspect of the algorithm because the objects were removed mostly from

Table 1. Birch3 Dataset Sampling

Log base	Radius	Objects	Objects %
-	-	100,000	100
4	50	44,098	44
4	100	24,745	28
4	200	14,835	15

**Fig. 1.** Birch dataset sampling

cluster centers while the border objects are preserved, and the densest areas are still preserved. The last sample (see Figure 1d) with the greatest radius shows that only the densest cluster centers are still preserved and other clusters are replaced by objects from the whole cluster. One very interesting fact is that even objects which are far from the center of any cluster and might be removed as noise are preserved if they form a cluster with the neighborhood.

4.2 Experiment 2: Properties of the Algorithm and Comparison with Random Sampling

The second experiment uses a dataset with two clusters in 2D space. The clusters were generated using Normal distribution with the following parameters: the first cluster contains 99000 objects, the mean is at $(0, 0)$ and the standard deviations are $(20, 20)$ along the horizontal and vertical axes. The second cluster contains 1000 objects with mean $(50, 50)$ and standard deviations $(2, 2)$. The clusters are not proportional and the smaller one contains only 1% of objects of the larger one. This should demonstrate the ability of the algorithm to preserve both clusters. Figures 2a - 2c depict the sampling results of the proposed algorithm. The original dataset is not visualized due to a lack of space, but may be seen as the light gray cluster. As can be seen, the algorithm preserves both clusters for any sampling rate. Moreover, the objects are more concentrated in the cluster centers. The setting of the algorithm parameters is depicted in Table 2.

The result of the unbiased random sampling algorithm is depicted in Figures 3a - 3c. The comparison with the proposed algorithm shows that the proposed

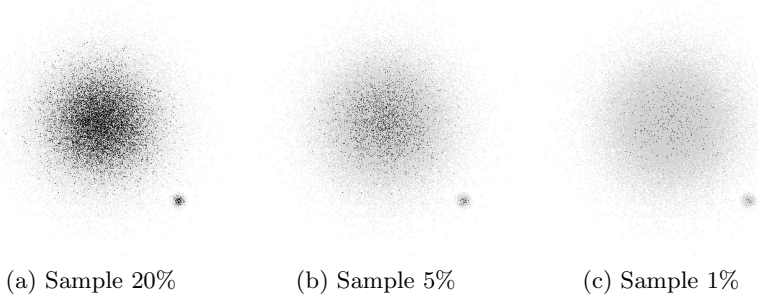


Fig. 2. Two-cluster dataset - Proposed algorithm

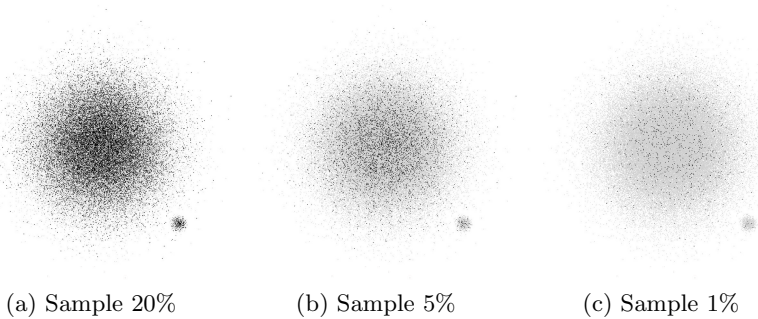


Fig. 3. Two-cluster dataset - Random samples

algorithm (Figures 2) is able to sample the data similarly to the random method, but closer inspection shows that the sample generated by the proposed algorithm is more concentrated in the cluster centers. Many objects in the random sample are located in the outer area of clusters and may be considered as noise, which may confuse further processing of the data.

Table 2. Two cluster Dataset Sampling

Log base	Radius	Objects	Objects %
-	-	100,000	100
2	0.0668	20,291	20
2	0.0300	5,196	5
2	0.0135	1,034	1

The samples in Figure 2 focused on mimicking the random samples with the emphasis on the cluster center. However, the different setting of the parameters may lead to completely different results. Figure 4 contains two 4% samples generated by the proposed algorithm. The first figure (4a) depicts the result similar

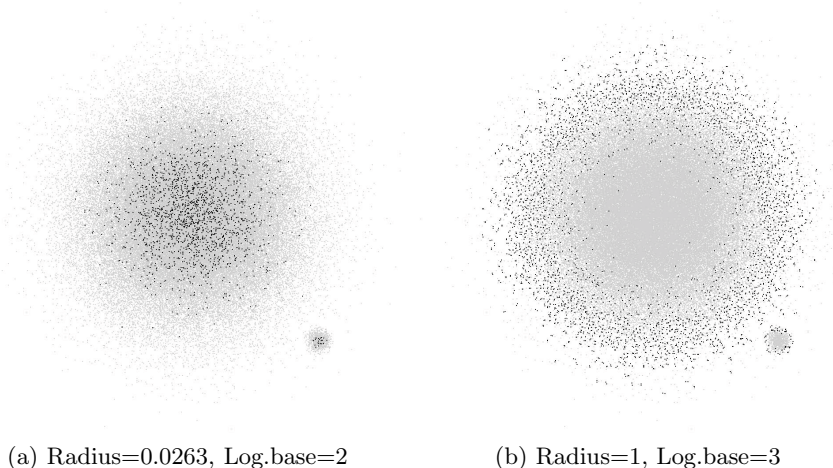


Fig. 4. Two-cluster dataset - Different parameters of proposed algorithm

to random sampling, while the second figure (4b) depicts the sample with objects concentrated on the border of the clusters.

4.3 Experiment 3: Large Real World Dataset

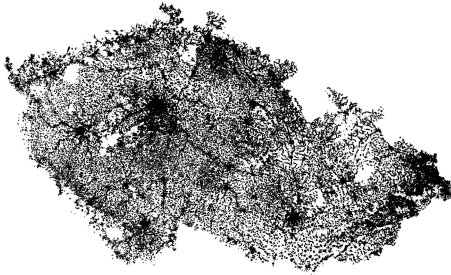
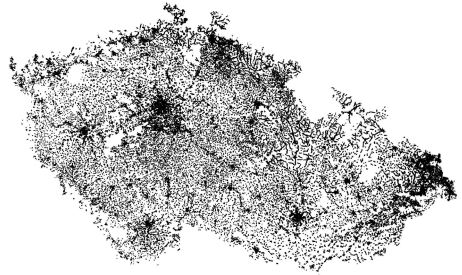
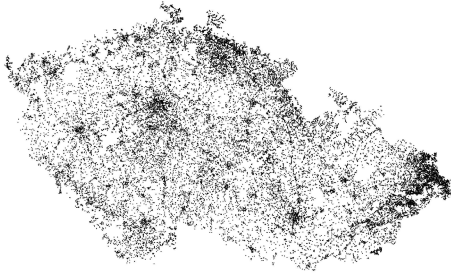
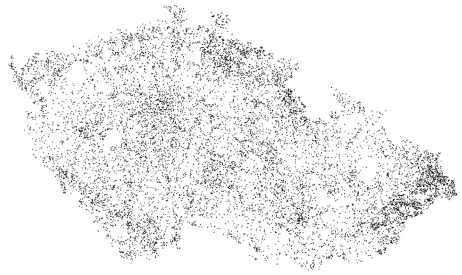
The third experiment uses a real-world 2-dimensional dataset which contains all address points in the Czech Republic provided by the government¹. This data set contains 2,7640,903 address points with coordinates in the S-JTSK coordination system (S-JTSK is a coord. system which has been used since the beginning of the 20th century in Czechoslovakia and the length unit is approximately one meter). The points are distributed more densely in the area of large cities, e.g. the densest place is in the middle of the image where the capital city Prague is located, but very dense areas are also found in the north and south, although the most populated area of the Czech Republic is in the east, where the Moravian-Silesian region is located. Similarly to the previous experiment, we discretize the distance with step 10. A summary of the data sampling with a different radius is depicted in Table 3; for the visualization, see Figures 5-8.

As can be seen from the figures, the visual comparison shows that the densest parts are still dense and recognizable even when a very large reduction is performed. Figure 7, where only 2% of points are preserved, clearly shows the largest cities in the Czech Republic and, moreover, shows that the eastern part of the republic has many densely populated places, therefore, it is the densest area preserved.

¹ <http://www.ruian.cz/> (in Czech).

Table 3. Czech Map Dataset Sampling

Log base Radius	Objects	Objects %
-	- 2,740,903	100
1.3	50 206,603	8
1.3	100 55,641	2
1.3	200 21,965	1

**Fig. 5.** Original dataset**Fig. 6.** Sample 8%**Fig. 7.** Sample 2%**Fig. 8.** Sample 1%

5 Conclusions

In this paper we presented our work in the field of sampling large-scale data. The approach is based on finding representatives in the input dataset. Measurement of representativeness is done by an analysis of the local properties and nearest neighbors. We consider the key features of the method to be its general applicability, natural scalability and flexibility.

We compared our algorithm with the unbiased random sampling method and performed tests on three different datasets. The results show that the algorithm is able to efficiently reduce the data using local information only.

Moreover, the experiments show that different parameters of the algorithm may lead to completely different results. One setting leads to behavior very similar to random sampling but with a greater focus on the cluster centers while a different setting leads to a sample with different point distribution - all sampled points are on the border of the clusters.

There are several more tasks to solve in future work, in particular, carrying out experiments on large-scale data and comparing them with other biased and unbiased methods. Since each dataset requires a different setting, it is necessary to make a deeper analysis of the dependencies between the parameters of the presented method, the processed dataset and the expected representative sample.

Acknowledgments. This work was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070), by the Development of human resources in research and development of latest soft computing methods and their application in practice project, reg. no. CZ.1.07/2.3.00/20.0072 funded by Operational Programme Education for Competitiveness, co-financed by ESF and state budget of the Czech Republic, and by SGS, VSB-Technical University of Ostrava, under the grant no. SP2014/110.

References

1. Barbar'a, D., DuMouchel, W., Faloutsos, C., Haas, P.J., Hellerstein, J.M., Ioannidis, Y., Jagadish, H., Johnson, T., Ng, R., Poosala, V., et al.: The new jersey data reduction report. In: IEEE Data Engineering Bulletin. Citeseer (1997)
2. Ernvall, J., Nevalainen, O.: An algorithm for unbiased random sampling. *The Computer Journal* 25(1), 45–47 (1982)
3. Kerdprasop, K., Kerdprasop, N., Sattayatham, P.: Weighted k-means for density-biased clustering. In: Tjoa, A.M., Trujillo, J. (eds.) *DaWaK 2005*. LNCS, vol. 3589, pp. 488–497. Springer, Heidelberg (2005)
4. Kivinen, J., Mannila, H.: The power of sampling in knowledge discovery. In: *Proceedings of the Thirteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, pp. 77–85. ACM (1994)
5. Kollios, G., Gunopulos, D., Koudas, N., Berchtold, S.: Efficient biased sampling for approximate clustering and outlier detection in large data sets. *IEEE Transactions on Knowledge and Data Engineering* 15(5), 1170–1187 (2003)
6. Manku, G.S., Rajagopalan, S., Lindsay, B.G.: Random sampling techniques for space efficient online computation of order statistics of large datasets. *ACM SIGMOD Record* 28, 251–262 (1999)
7. Nanopoulos, A., Manolopoulos, Y., Theodoridis, Y.: An efficient and effective algorithm for density biased sampling. In: *Proceedings of the Eleventh International Conference on Information and Knowledge Management*, pp. 398–404. ACM (2002)
8. Palmer, C.R., Faloutsos, C.: *Density biased sampling: an improved method for data mining and clustering*, vol. 29. ACM (2000)
9. Toivonen, H., et al.: Sampling large databases for association rules. In: *VLDB*, vol. 96, pp. 134–145 (1996)

10. Vitter, J.S.: Random sampling with a reservoir. *ACM Transactions on Mathematical Software (TOMS)* 11(1), 37–57 (1985)
11. Vitter, J.S.: Faster methods for random sampling. *Communications of the ACM* 27(7), 703–718 (1984)
12. Zhang, T., Ramakrishnan, R., Livny, M.: Birch: A new data clustering algorithm and its applications. *Data Mining and Knowledge Discovery* 1(2), 141–182 (1997)
13. Zhou, S., Zhou, A., Cao, J., Wen, J., Fan, Y., Hu, Y.: Combining sampling technique with dbscan algorithm for clustering large spatial databases. In: Terano, T., Liu, H., Chen, A.L.P. (eds.) *PAKDD 2000*. LNCS, vol. 1805, pp. 169–172. Springer, Heidelberg (2000)

Diagonal Interacting Multiple Model H_∞ Filtering for Simultaneous Sensor Localization and Target Tracking with NLOS Mitigation^{*}

Xiaoyan Fu², Yuanyuan Shang¹, Hui Ding², and Xiuzhuang Zhou¹

¹ Beijing Engineering Research Center of High Reliable Embedded System,

² Beijing Key Laboratory of Electronic System Reliability Technology, College of Information Engineering, Capital Normal University, Beijing
fuxiaosg@163.com

Abstract. This paper is devoted to the problem of simultaneous localization and tracking (SLAT) in non-line-of-sight (NLOS) environments. By combining a target state and a sensor node location into an augmented vector, a discrete-time stochastic systems with Markov jump parameters is used to describe the switching of LOS/NLOS. A robust algorithm—diagonal interacting multiple model algorithm based on H_∞ filtering (DIMMH) is presented for simultaneous refinement of sensors' positions and target tracking when measurement noise is of unknown statistics. We use a measurement model from a real mine to handle all non-Gaussian uncertainties typical for mining environments, and analyze the performance of the classical interacting multiple model (IMM) algorithm, the DIMM algorithm and the cubature Kalman filter (CKF).

1 Introduction

Mine tunnels are extensive labyrinths with irregularly-shaped walls, in which a hundreds of employees are working on extraction of valuable ores and minerals. The miners work under hazardous environmental conditions caused by the high humidity and poor ventilation, the presence of flammable and toxic gases, corrosive water and dust, and the dangers of rock falls and mine collapses [1]-[3]. The knowledge of the last location of the miners is especially important in the aftermath of the accidents such as mine collapse or explosion, but can be also used for task optimization and traffic management. A GPS-based localization system provides the global position of a mobile vehicle or object in outdoor environment [4]. However, the GPS-based system has an inherent disadvantage because the GPS signal cannot be available in indoor scenarios [5]. A wireless sensor network (WSN) can be deployed across the mine to monitor the environmental conditions such as stability, humidity and toxic gas levels. The information obtained

^{*} This work was supported by the National Natural Science Foundation of China (11178017, 61373090, 61303104, 61203238) and the Beijing Natural Science Foundation of China (4132014).

from the sensors can be used to control the ventilation system, and determine the unsafe areas and rescue paths. Beyond this ability, a WSN can be used to track the personnel, mobile equipment and vehicles [1].

However, the state-of-the-art algorithms [3],[6],[7] assume that the positions of the sensors are perfectly known, which is not necessarily true due to the imprecise placement and/or sensor drops caused by vibrations or wall collapses¹. Though the miners can periodically verify if all the sensors' positions are correct, this approach is too costly and even infeasible in some areas due to the on-going mining activities. An effective option is to let the sensors estimate their individual positions while tracking a target in mine tunnels. In [9], the problem of target tracking by a network with unknown sensor positions has been addressed, which is also defined as simultaneous localization and tracking (SLAT). In [10], by assuming that sensors are randomly deployed, a sequential quasi-Monte Carlo-based filter has been developed to address the problem of SLAT. A distributed variational filter for SLAT has been proposed in [11], in which the energy consumption and bandwidth consumption are considered. Although much work has been done to SLAT, as shown in [9]-[11], almost all the proposed filters are derived based on the Sequence Monte Carlo (SMC) method, which are also known to be of high computational costs. Moreover, the received signal strength model is used to generate measurements in the aforementioned literature, whereas the non-line-of-sight (NLOS) effect is not considered.

In fact, there might be no direct path between a target and a sensor in a mine tunnels environment which are extensive labyrinths with irregularly-shaped walls. Furthermore, the propagating signal may travel excess path lengths of hundreds of meters due to reflection and diffraction. This error is referred to the NLOS error and may yield an estimation bias if not be addressed. To mitigate the NLOS error, many strategies have been proposed, a two state Markov process has been employed to describe the transition of the LOS/NLOS, and an interacting multiple-model (IMM) approach is used to derive the target-state estimate in [12]. Further improved results have been obtained in [13]-[15]. It is noted that combining the state estimations and corresponding covariance according to the scalar weights in the IMM algorithm. But in the problem of SLAT, the state augmented vector is the combining a target state and a sensor node location, The probability distribution of target state and sensor node location is difference, IMM algorithm can not distinguish the effects produced by different dimensions of the state. Moreover, simultaneous sensor localization and target tracking in a mine tunnel, the measurement noise is of unknown statistics.

In this paper, H_∞ filtering are introduced into DIMM algorithm for SLAT in mine tunnels. We choose H_∞ filtering to deal with the state estimate problem in view of the following advantages of H_∞ estimate [16]: 1) H_∞ filtering provides a rigorous method for dealing with systems that have model uncertainty. 2) H_∞ filtering can be used to guarantee stability margins or minimize the worst

¹ Although not available in mines nowadays, we also envision that the uncertain sensors' positions can be an outcome of some (cooperative) sensor network localization algorithm[1],[8].

case estimate error. 3) H_∞ filtering may be more appropriate for systems whose models change unpredictably and when it is too complex or time consuming to model identification or gain scheduling. H_∞ filtering can deal with arbitrary signals with only a requirement of bounded noise, which replaces the Kalman filter method of modeling the noise as a random process. The results of H_∞ estimate are more robust than that in the signal models with uncertain parameters. In the DIMM algorithm, the diagonal matrices from the optimal multi-model fusion criterion are used as the weights of models. distinguish the effects produced by different dimensions of the state. The original edition of the DIMM algorithm can be found in our previous conference paper [17].

This paper is organized as follows. In section II, the problem of SLAT in NLOS environments is formulated as state estimate of discrete-time stochastic systems with Markov switching parameters, and IMM algorithm is reviewed and analyzed, which provides preliminaries for the following sections. In Section III, diagonal interacting multiple model algorithm based on H_∞ filtering (DIMMH) is presented. The conclusions are provided in Section IV.

2 Preliminaries

2.1 Markov Jump Systems Tracking Problem

Consider the following Markov jump system:

$$\mathbf{x}(k+1) = \mathcal{F}\mathbf{x}(k) + \mathcal{T}\nu(k) \quad (1)$$

$$\tau(k) = g_j(\mathbf{x})(k) + \omega_j(k) \quad (2)$$

where the state vector $\mathbf{x}(k)$ is an n -dimensional vector, the observation process $\mathbf{z}(k)$ is an m -dimensional vector, and the subscript $j \in \mathbb{S} = \{1, 2\}$ denotes the model. The matrix functions $\mathcal{F}(\cdot)$, $\mathcal{T}(\cdot)$ and $g_j(\cdot)$ are known. The model-dependent process noise is assumed to be a Gaussian random process with:

$$E[\nu(k)] = 0, \quad E[\nu(k)\nu(k)^T] = Q_j \quad (3)$$

The measurement model switch between two types of the LOS and the NLOS situations. Then, we formulate the problem of mobile location estimation into the framework of nonlinear filtering for jump Markov systems with unknown statistics noise. Without loss of generality, exogenous inputs $D\mathbf{u}(k)$ can be considered in (1), but for notational convenience, here they are omitted.

– LOS case

$$\tau(k) = \frac{2\|x(k) - z(k)\|}{c} + \tau_{PT} + \omega^q(k) + \omega^m(k) \quad (4)$$

$$W_{t,n}^q \sim p_q(\omega_q) = \text{Unif}(\omega_q; 0, \frac{2D\sqrt{3}}{c}), W_{t,n}^m \sim p_m(\omega_m) = \mathcal{N}(\omega; 0, \sigma_w^2)$$

– NLOS case

$$\tau_{t,n} = \frac{2\|x_t - z(k)\|}{c} + \tau_{PT} + \omega^q(k) + \omega^m(k) \quad (5)$$

$$\omega^q(k) \sim p_q(\omega_q) = \text{Unif}(\omega_q; 0, \frac{2D\sqrt{3}}{c}), \omega^m(k) \sim p_m(\omega_m) = \mathcal{B}(\omega; \mu_w, \alpha_w, \gamma_w)$$

where τ_{PT} is a known processing time on a target found by calibration, $c = 3 \cdot 10^8 \text{ m/s}$ is the speed of light, $\omega^q(k)$ is quantization noise, and $\omega^m(k)$ is measurement noise. Note that the quantization noise is written outside the norm using an upper bound of the triangle inequality (i.e., $\|a + b\| \leq \|a\| + \|b\|$), which represents the worst case scenario. where σ_w is the standard deviation of the LOS component of the noise, and $B(\cdot)$ is a Weibull distribution with scale α_w , shape γ_w , and location parameters μ_w ($\alpha_w > 0$, $\gamma_w > 0$, $\omega > \mu_w$),

Let M_j^k denotes the flight model j at time k . The model dynamics are modeled as a finite Markov chain with known model-transitions probabilities from model i at time $k - 1$ to model j at time k [18], [19].

$$\pi_{ij} \triangleq \text{Prob}\{M_j^k \mid M_i^{k-1}\} = \mathbf{P}\{M_j^k \mid M_i^{k-1}\} \quad (6)$$

$$0 \leq \pi_{ij} \leq 1, \quad \sum_{j=1}^s \pi_{ij} = 1, \quad i, j \in \mathbb{S} \quad (7)$$

The initial state distribution of the Markov chain is $\varphi = [\varphi_1, \dots, \varphi_s]$, where

$$0 \leq \varphi_j \leq 1, \quad \sum_{j=1}^s \varphi_j = 1, \quad j \in \mathbb{S} \quad (8)$$

This Markov chain description of the target's models is used to model the unknown inputs.

It is also possible to use UWB and wideband received-signal strength (RSS) measurements using the models in [20], respectively. The noise in that case is a mixture of two Gaussians, corresponding to LOS and NLOS, respectively. However, RSS can only provide coarse distance estimates since it cannot exploit the very large bandwidth of the signal [1].

2.2 IMM Algorithm

IMM algorithm is the most prevalent for the state estimate of discrete-time stochastic systems with Markov switching parameters. The following steps are associated with IMM algorithm [21]:

Step 1. Calculate the mixed initial probability for the filter matched to model M_j^k ($j \in \mathbb{S}$)

Step 2. Calculate the mixed initial state and corresponding covariance for the filter matched to model M_j^k

Step 3. Kalman Filtering

Step 4. Combine the state estimates and corresponding covariances according to the updated weights

Remark 1. In IMM algorithm, updated weights of models are derived from the hybrid of *pdfs* and probability masses. It is known that any probability mass must be a value in the interval $[0, 1]$, but any *pdf* has no such restriction, thus, the two kinds of values are at different levels. The resulting outcome μ_j^k is just an approximate probability. Moreover, when the measurement noise is of unknown statistics, IMM algorithm will produce more error. It is therefore necessary to propose an optimal filtering approach for the state estimate with uncertain noise.

3 Diagonal Interacting Multiple Model Algorithm Based On H_∞ Filtering

3.1 H_∞ Filtering

Consider the systems in (1-2) in the case where the process noise ν and the measurement noise ω_k are assumed to be energy bounded l_2 signals whose statistical properties are unknown.

Unlike the Kalman filter which aims to give the minimum mean-square estimate of the state vector \mathbf{x}_k , the optimal H_∞ filter tries to obtain the arbitrary linear combination of the state \mathbf{x}_k using the measurements \mathbf{Y}_k such that the effect of the worst disturbance on the estimate error is minimized, namely, $\mathbf{z}_k = L_k \mathbf{x}_k$ where L_k is a known matrix. Here, we are interested in state estimate, so L_k is taken as an identity matrix I . Let $\hat{\mathbf{x}}_{k|k}$ denote the estimate of \mathbf{x}_k given measurements \mathbf{Y}_k , and the estimate error is denoted as $\mathbf{e}_k = \hat{\mathbf{x}}_{k|k} - \mathbf{x}_k$.

3.2 Cubature Kalman Filters

Consider the filtering problem of nonlinear dynamic system (1-2) with additive noise.

It is known that the Bayesian filter is rendered tractable when all conditional densities are assumed to be Gaussian. In this case, the Bayesian filter solution reduces to computing multi-dimensional integrals, whose integrands are all of the form *nonlinear function* \times *Gaussian*. The CKF exploits the properties of highly efficient numerical integration methods known as cubature rules for those multi-dimensional integrals [22]. Moreover, The CKF is numerically accurate and easily extendable to high-dimensional problems. In this paper, we extend the CKF and H_∞ filtering to form a cubature H_∞ filtering. The cubature H_∞ filtering is not only useful for multi-state estimation but it can also handle nonlinear and non-Gaussian systems.

3.3 DIMMH Algorithm

In this section, cubature H_∞ filtering is induced to receive the state estimate instead of the Kalman filter to obtain the optimal state estimates when the noise with unknown statistics. The following steps are associated with the DIMMH algorithm.

Step 1. Calculate the mixed initial diagonal-matrix-weight for the filter matched to model M_j^k ($j \in \mathbb{S}$):

$$\begin{aligned}
 B_{i|j}(k|k) &\triangleq \mathbf{P}\{M_i^{k-1}|M_j^k, Z^{k-1}\} \\
 &= \frac{\pi_{ij}B_i^{k-1}}{\sum_{i=1}^s \pi_{ij}B_i^{k-1}} \\
 &= \begin{pmatrix} \frac{\pi_{ij}b_{i1}}{\sum_{i=1}^s \pi_{ij}b_{i1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\pi_{ij}b_{in}}{\sum_{i=1}^s \pi_{ij}b_{in}} \end{pmatrix} \tag{9}
 \end{aligned}$$

where

$$\begin{aligned}
 B_i^{k-1} &= \text{diag}(b_{i1}, b_{i2}, \dots, b_{in}) \\
 &\triangleq \mathbf{P}\{M_i^{k-1}|Z^{k-1}\} \tag{10}
 \end{aligned}$$

Step 2. Calculate the mixed initial state and corresponding covariance for the filter matched to model $M_j(k)$ ($j \in \mathbb{S}$):

$$\hat{\mathbf{x}}_{0j}(k|k) = \sum_{i=1}^s B_{i|j}(k|k)\hat{\mathbf{x}}_i^{k-1} \tag{11}$$

$$\begin{aligned}
 P_{0j}(k|k) &= \sum_{i=1}^s B_{i|j}(k|k)\{P_i^{k-1} + [\hat{\mathbf{x}}_i^{k-1} - \hat{\mathbf{x}}_{0j}(k|k)] \\
 &\quad \times [\hat{\mathbf{x}}_i^{k-1} - \hat{\mathbf{x}}_{0j}(k|k)]^T\} \tag{12}
 \end{aligned}$$

Step 3. Cubature H_∞ filtering ($j \in \mathbb{S}$)

$$\hat{\mathbf{x}}_{k|k-1} = \frac{1}{2n} \sum_{i=1}^{2n} \chi_{i,k|k-1}^* \tag{13}$$

$$P_{k|k-1} = \frac{1}{2n} \sum_{i=1}^{2n} \chi_{i,k|k-1}^{*T} - \hat{\mathbf{x}}_{k|k-1} \hat{\mathbf{x}}_{k|k-1}^T + \mathcal{T} \tilde{Q} \mathcal{T}^T \tag{14}$$

$$\hat{\mathbf{x}}_j^k = \hat{\mathbf{x}}_{k|k-1} + K_k(\tau_k - \hat{\tau}_{k|k-1}) \tag{15}$$

$$P_j^k = P_{k|k-1} - K_k P_{\tau\tau, k|k-1} K_k^T - \gamma^- 2I_n \tag{16}$$

$$K_k = P_{x\tau, k|k-1} P_{\tau\tau, k|k-1}^{-1} \tag{17}$$

where

$$\chi_{i,k|k-1}^* = \mathcal{F}\chi_{i,k-1|k-1} \quad (18)$$

$$\tau_{i,k|k-1} = g(\chi_{i,k|k-1}) \quad (19)$$

$$\chi_{i,k|k-1} = \sqrt{P_{k|k-1}}\xi_i + \hat{\mathbf{x}}_{k|k-1} \quad (20)$$

$$\hat{\tau}_{k|k-1} = \frac{1}{2n} \sum_{i=1}^{2n} \tau_{i,k|k-1} \quad (21)$$

$$P_{\tau\tau,k|k-1} = \frac{1}{2n} \sum_{i=1}^{2n} \tau_{i,k|k-1} \tau_{i,k|k-1}^T - \hat{\tau}_{k|k-1} \hat{\tau}_{k|k-1}^T \quad (22)$$

$$P_{x\tau,k|k-1} = \frac{1}{2n} \sum_{i=1}^{2n} \chi_{i,k|k-1} \tau_{i,k|k-1}^T - \hat{x}_{k|k-1} \hat{\tau}_{k|k-1}^T \quad (23)$$

Step 4. Combine of the state estimates and corresponding covariances according to the updated diagonal-matrix-weight:

$$\hat{\mathbf{x}}_D(k) = \sum_{j=1}^s B_j^k \hat{\mathbf{x}}_j^k \quad (24)$$

Updated diagonal-matrix-weight of model M_j^k is

$$B_j^k = \text{diag}(b_{j1}, b_{j2}, \dots, b_{jn}) \quad (25)$$

where

$$[b_{1i}, b_{2i}, \dots, b_{si}] = \frac{\mathbf{e}^T(\mathcal{P}^i)^{-1}}{\mathbf{e}^T(\mathcal{P}^i)^{-1}\mathbf{e}} \quad (26)$$

with

$$\mathbf{e} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{s \times 1}, \quad \mathcal{P}^i = \begin{bmatrix} P_1^{(ii)} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & P_s^{(ii)} \end{bmatrix} \quad (27)$$

and $P_j^{(ii)}$ is the i th diagonal element of matrix P_j ($P_j = E[\tilde{\mathbf{x}}_j \tilde{\mathbf{x}}_j^T]$).

The error variance matrix of the optimal fusion estimate is

$$P_D(k) = \text{diag}[P_{D1}, P_{D2}, \dots, P_{Dn}] \quad (28)$$

where

$$P_{Di} = [\mathbf{e}^T(\mathcal{P}^i)^{-1}\mathbf{e}]^{-1} \quad (29)$$

Remark 2. For solving the problem of SLAT, the state of target and a sensor node location are combined into an augmented vector. The noise statistics is different between tracked target and sensor node. In DIMMH algorithm, the diagonal matrices from the optimal multi-model fusion criterion are used as the weights of models, which can be viewed as the joint probabilities of models. That is to say, the state vector is segmented into n scalars to carry on estimating, and every element of diagonal matrix can be interpreted as a probability mass of the model with dimension one. The new algorithm can not only avoid the mixture of likelihood function and probability mass and distinguish the effects produced by different dimensions of the state like DIMM algorithm but also deal with the noise with unknown statistics.

Remark 3. Another difference between the proposed algorithm and the celebrated IMM estimator lies on the fact that the H_∞ filtering and cubature rule are combined. The cubature rule is employed to deal with the nonlinear measurements in this work which is a derivative-free approximation scheme.

It is interesting to note that H_∞ filtering has the same observer structure as that of the Kalman filter, and \tilde{Q} and \tilde{R} play the same role as the variances of the process noise and the measurement noise when using the Kalman filtering [23]. Indeed, the H_∞ filter is equivalent to the Kalman filter in the Krein space and the H_∞ filter exists if and only if $P_k^{-1} > 0$ [24]. Specifically, the H_∞ filter is reduced to the Kalman filter when $\gamma \rightarrow \infty$. Thus, the γ may be thought as a tuning parameter to control the tradeoff between H_∞ performance and minimum variance performance. The optimal H_∞ filter can also be interpreted in the frequency domain as an estimate that minimizes the peak error power whereas the Kalman filter aims to minimize the average error power or error covariance.

4 Conclusions and Future Work

In the paper, DIMMH algorithm is presented for maneuvering target tracking. It is principally similar to the popular IMM algorithm and DIMM algorithm proposed in our previous paper. The difference lies in the use of filtering. To obtain the optimal state estimates in the nonlinear switching system when the noise with unknown statistics, H_∞ filtering and cubature rule are combined instead of the Kalman filter. In future work, we will research on how to deal with arbitrary uncertain noise stretching beyond l_2 signal and demonstrate the computer simulations for indicate the superiority of proposed algorithms.

References

1. Savic, V., Wymeersch, H., Larsson, E.G.: Simultaneous sensor localization and target tracking in mine tunnels. In: IEEE Int. Conf. Information Fusion, pp. 1427–1433 (July 2013)

2. Misra, P., Kanhere, S., Ostry, D., Jha, S.: Safety assurance and rescue communication systems in high-stress environments: A mining case study. *IEEE Communications Magazine* 48, 66–73 (2010)
3. Chehri, A., Fortier, P., Tardif, P.M.: UWB-based sensor networks for localization in mining environments. *Ad Hoc Networks* 7, 987–1000 (2009)
4. Ahn, H.S., Ko, K.H.: Simple pedestrian localization algorithms based on distributed wireless sensor networks. *IEEE Trans. Ind. Electron.* 56(10), 4296–4302 (2009)
5. Hur, H., Ahn, H.S.: Discrete-time H_∞ filtering for mobile robot localization using wireless sensor network. *IEEE Sensors Journal* 13(1), 245–252 (2013)
6. Dayekh, S., Affes, S., Kandil, N., Nerguizian, C.: Cooperative localization in mines using fingerprinting and neural networks. In: *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6 (April 2010)
7. Li, M., Liu, Y.: Underground coal mine monitoring with wireless sensor networks. *ACM Trans. Sensor Networks* 5, 1–29 (2009)
8. Wymeersch, H., Penna, F., Savic, V.: Uniformly reweighted belief propagation for estimation and detection in wireless networks. *IEEE Transactions on Wireless Communications* 11, 1587–1595 (2012)
9. Taylor, C., Rahimi, A., Bachrach, J., Shrobe, H., Grue, A.: Simultaneous localization, calibration, and tracking in an ad hoc sensor network. In: *Proc. 5th Int. Conf. Inf. Process. Sensor Netw.*, pp. 27–33 (2006)
10. Aggarwal, P., Wang, X.: Joint sensor localisation and target tracking in sensor networks. *IET Radar, Sonar Navig.* 5(3), 225–233 (2011)
11. Teng, J., Snoussi, H., Richard, C., Zhou, R.: Distributed variational filtering for simultaneous sensor localization and target tracking in wireless sensor networks. *IEEE Trans. Veh. Technol.* 61(5), 2305–2318 (2012)
12. Liao, J.F., Chen, B.S.: Robust mobile location estimator with NLOS mitigation using interacting multiple model algorithm. *IEEE Trans. Wireless Commun.* 5(11), 3002–3006 (2006)
13. Yang, C.Y., Chen, B.S., Liao, F.K.: Mobile location estimation using fuzzy-based IMM and data fusion. *IEEE Trans. Mobile Comput.* 9(10), 1424–1436 (2010)
14. Hammes, U., Zoubir, A.M.: Robust MT tracking based on M-estimation and interacting multiple model algorithm. *IEEE Trans. Signal Process.* 59(7), 3398–3409 (2011)
15. Li, W., Jia, Y., Du, J., Zhang, J.: Distributed Multiple-Model Estimation for Simultaneous Localization and Tracking With NLOS Mitigation. *IEEE Transactions on Vehicular Technology* 62(6) (July 2013)
16. Grimble, M., Johnson, M.: H_∞ robust control design—A tutorial review. *Computing and Control Engineering Journal* 6, 275–282 (1991)
17. Fu, X., Jia, Y., Du, J., Yuan, S.: New interacting multiple model algorithms for tracking of maneuvering target. *IET Control Theory & Applications* 4, 2184–2194 (2010)
18. Blackman, S.S., Popoli, R.F.: *Design and analysis of modern tracking systems*. Artech House, Boston (1999)
19. Yepes, J.L., Hwang, I., Rotea, M.: New algorithms for aircraft intent inference and trajectory prediction. *AIAA Journal of Guidance, Control, and Dynamics* 30(2), 370–382 (2007)
20. Chehri, A., Fortier, P., Tardif, P.M.: Characterization of the ultra-wideband channel in confined environments with diffracting rough surfaces. *Wireless Personal Communications (Springer)* 62, 859–877 (2012)

21. Li, X.R., Jilkov, V.P.: Survey of maneuvering target tracking. Part V. Multiple-model methods. *IEEE Transactions on Aerospace and Electronic Systems* 41(4), 1255–1321 (2005)
22. Arasaratnam, I., Haykin, S.: Cubature Kalman filters. *IEEE Transactions on Automatic Control* 54(6), 1254–1269
23. Fu, X., Jia, Y., Du, J., Yuan, S.: A novel interacting multiple model algorithm based on multi-sensor optimal information fusion rules. In: *American Control Conference* (2009)
24. Simon, D.: Kalman filtering with state constraints: A survey of linear and nonlinear algorithms. *IET Control Theory and Application* 8, 1303–1318 (2010)

Part II
**Intelligent Data Analysis and
Its Applications**

A Projection-Based Approach for Mining Highly Coherent Association Rules

Chun-Hao Chen¹, Guo-Cheng Lan², Tzung-Pei Hong^{2,4,*},
Shyue-Liang Wang³, and Yui-Kai Lin⁴

¹ Department of Computer Science and Information Engineering,
Tamkang University, Taipei, 251, Taiwan

² Department of Computer Science and Information Engineering,
National University of Kaohsiung, Kaohsiung, 811, Taiwan

³ Department of Information Management,
National University of Kaohsiung, Kaohsiung, 811, Taiwan

⁴ Department of Computer Science and Engineering,

National Sun Yat-sen University, Kaohsiung, 804, Taiwan

chchen@mail.tku.edu.tw, rrfuheiy@gmail.com,

{tphong, slwang}@nuk.edu.tw, m993040046@sudent.nsysu.edu.tw

Abstract. In our previous approach, we proposed an apriori-based algorithm for mining highly coherent association rules, and it is time-consuming. In this paper, we present an efficient mining approach, which is a projection-based technique, to speed up the execution of finding highly coherent association rules. In particular, an indexing mechanism is designed to help find relevant transactions quickly from a set of data, and a pruning strategy is proposed as well to prune unpromising candidate itemsets early in mining. The experimental results show that the proposed algorithm outperforms the traditional mining approach for a real dataset.

Keywords: Data mining, propositional logic, highly coherent rule, index mechanism, pruning strategy.

1 Introduction

The association rule mining is one of the well-known techniques for this purpose. Agarwal et al. proposed Apriori approach for association mining [2]. Lots of various algorithms based on this approach have been proposed for mining association rules [1, 3, 4, 5]. In those approaches, various types of association rule mining algorithm, such as weighted association rule mining [3, 10] and fuzzy association rules mining [7], have been proposed according to the specified goals.

Although these association rule mining approaches derive a lot of rules, they are common sense, which also making users difficult to use, especially for business applications, and may produce misleading description rules. For example, a rule like

* Corresponding author.

“If bread is bought, then milk is bought” may be misleading because customers that buy bread may not buy milk. To solve this problem, Sim et al. proposed an algorithm for mining coherent rules based on the properties of propositional logic without a minimum support threshold. In their approach, if a rule satisfies the logic equivalence, then it is a coherent rule. In other words, in that approach, if the rule $X \rightarrow Y$ exists, then the rule $\neg X \rightarrow \neg Y$ also exists.

Chen et al. proposed an apriori-based algorithm for mining highly coherent association rules with consideration of support measure [6]. Since the execution time of Chen's approach is time-consuming, this study proposes a projection-based approach, named the projection-based coherent rule mining algorithm (PCMA), to speed up the process for mining highly coherent association rules from the given transaction database. Since generating candidate coherent itemsets is time-consuming, the proposed approach first calculates the corresponding intervals of lower and upper bounds of sub-itemsets in consequent part of a subsequence itemset. These intervals are then used for removing itemsets that cannot become highly coherent itemsets to speed up the mining process. The contingency tables of the valid itemsets are calculated and used for checking whether they satisfy the conditions of logical equivalence. If yes, then the itemsets are used to generate highly coherent association rules. The proposed approach is adopted the idea of projection, which avoids generating unnecessary candidates and reduce the times of scanning databases. The efficiency of process can be improved. Experiments made on two datasets show that the proposed approach is efficient.

2 Preliminaries

One of the main issues of association rule mining is how to define the appropriate minimum support and minimum confidence. Although an appropriate minimum support may exist, it is difficult to find it. Numerous studies have proposed method for finding an appropriate minimum support [8, 11]. The coherent rule mining algorithm, proposed by Sim et al., is based on the properties of propositional logic [9]. In the approach, by using the properties of propositional logic, relationship between items can be derived directly without knowing the appropriate value of the minimum support. The main concept of the approach maps the association rules to equivalences. Each mapping from an association rule to an equivalence should satisfy conditions, which are shown in Table 1.

Table 1. Four conditions for mapping rules to equivalence

Equivalences	$p \equiv q$	$\neg p \equiv \neg q$
Association Rules	$X \rightarrow Y$	$\neg X \rightarrow \neg Y$

True or False on Association Rules	Required Conditions	
T	$X \rightarrow Y$	$\neg X \rightarrow \neg Y$
F	$X \rightarrow \neg Y$	$\neg X \rightarrow Y$
F	$\neg X \rightarrow Y$	$X \rightarrow \neg Y$
T	$\neg X \rightarrow \neg Y$	$X \rightarrow Y$

In Table 1, X and Y are two itemsets. Association rule $X \rightarrow Y$ is mapped to $p \equiv q$ if and only if (1) $X \rightarrow Y$ is true; (2) $\neg X \rightarrow Y$ is false; (3) $X \rightarrow \neg Y$ is false; and (4) $\neg X \rightarrow \neg Y$ is true. When used in multiple transactions, association rules can be mapped to implications as follows: $X \rightarrow Y$ is mapped to implication $p \rightarrow q$ if and only if (1) $\text{Sup}(X, Y) > \text{Sup}(X, \neg Y)$; (2) $\text{Sup}(X, Y) > \text{Sup}(\neg X, Y)$; (3) $\text{Sup}(X, Y) > \text{Sup}(X, \neg Y)$; and (4) $\text{Sup}(X, Y) > \text{Sup}(\neg X, \neg Y)$. In the same way, other association rules can be mapped to implications based on comparisons between supports (called *pseudoimplications*). Sim et al. extended *pseudoimplications* to coherent rules. The following four conditions must be satisfied for a coherent rule: (1) $\text{Sup}(X, Y) > \text{Sup}(\neg X, Y)$; (2) $\text{Sup}(X, Y) > \text{Sup}(X, \neg Y)$; (3) $\text{Sup}(\neg X, \neg Y) > \text{Sup}(\neg X, Y)$; and (4) $\text{Sup}(\neg X, \neg Y) > \text{Sup}(X, \neg Y)$. These four conditions are represented as a contingency table in Table 2.

Table 2. Contingency table of a rule

Frequency of co-occurrences		Consequence Y	
		Y	$\neg Y$
Antecedent X	X	$Q_1 = \text{Sup}(X, Y)$	$Q_2 = \text{Sup}(X, \neg Y)$
	$\neg X$	$Q_3 = \text{Sup}(\neg X, Y)$	$Q_4 = \text{Sup}(\neg X, \neg Y)$

The concept of a coherent rule is that if a rule $X \rightarrow Y$ exists, then the rule $\neg X \rightarrow \neg Y$ should also exist. By utilizing this concept, to the present study proposes an association rule mining algorithm that uses the four conditions in the mining process for deriving highly coherent association rules from transactions.

3 Problem Statement and Definitions

For the formal definition of highly coherent association rule mining, a set of terms related to highly coherent association rule mining with consideration of positive and negative relationships of items is defined as follows.

Definition 1. An itemset X is a subset of items, $X \subseteq I$. If $|X| = r$, the itemset X is called an r -itemset. Here $I = \{i_1, i_2, \dots, i_m\}$ is a set of items, which may appear in transactions. For example, the itemset $\{i_4 i_6\}$ contains 2 items and is called a 2-itemset.

Definition 2. A transaction (*Trans*) is composed of a set of purchased items.

Definition 3. A transaction database D is composed of a set of transactions. That is, $D = \{Trans_1, Trans_2, \dots, Trans_y, \dots, Trans_z\}$, where $Trans_y$ is the y -th transaction in D .

Definition 4. The support sup_i of an item i in D is the number of transactions including the item i in D over the number of all transactions in D .

Definition 5. The support sup_X of an itemset X in D is the number of transactions including the item X in D over the number of all transactions in D .

Definition 6. Let α be a pre-defined minimum support threshold. An itemset X is called a frequent itemset (abbreviated as *FI*) if $sup_X \geq \alpha$.

Based on the above definitions, a set of frequent itemsets could first be found in a transaction database. After the process of finding frequent itemsets, all highly coherent itemsets can be found in the set of frequent itemsets, and then highly coherent rules can also be generated from the set of highly coherent itemsets. Accordingly, several terms related to coherent rule mining are stated below.

Definition 7. Let S be a frequent q -itemset with items (i_1, i_2, \dots, i_q) , $q \geq 2$. All possible pairs of sub-itemsets X and Y from the itemset S , where X and Y are subsets of S , $(X \cup Y) = S$, and $(X \cap Y) = \emptyset$, are first generated. For each pair (X, Y) , the four kinds of support values for antecedent X and consequent Y in the contingency table (as shown in Table 4) by using all frequent itemsets (L_{all}) could be defined as Q_1 : Sup_{XY} , Q_2 : $Sup_{X \rightarrow Y}$, Q_3 : $Sup_{\neg X \rightarrow Y}$ and Q_4 : $Sup_{\neg X \rightarrow \neg Y}$, respectively. If the four kinds of support values of each pair (X, Y) meet the four conditions, which are $Q_1 > Q_2$, $Q_1 > Q_3$, $Q_4 > Q_2$, and $Q_4 > Q_3$, the itemset S is a highly coherent itemset, *HC*. Note that Q_2 , Q_3 and Q_4 can be calculated by $(Sup_X - Q_1)$, $(Sup_Y - Q_1)$, and $(T - Q_1 - Q_2 - Q_3)$, respectively. Note that the two symbols, X and $\neg X$, in Table 3 represent the transactions with and without X in a set of transactions, respectively.

Table 3. The contingency table

Frequency of concurrences		Consequence, Y	
		Y	$\neg Y$
Antecedent, X	X	Q_1	Q_2
	$\neg X$	Q_3	Q_4

Definition 8. Let X be a highly coherent q -itemset with items (x_1, x_2, \dots, x_q) , $q \geq 2$. Then, the confidence $conf_R$ of an association rule in D , which is denoted as $\{i_1 \wedge \dots \wedge i_{k-1} \wedge i_{k+1} \wedge \dots \wedge i_q\} \rightarrow \{i_k\}$, is defined as follows:

$$conf_{\{i_1 \wedge \dots \wedge i_{k-1} \wedge i_{k+1} \wedge \dots \wedge i_q\} \rightarrow \{i_k\}} = \frac{\sup_X}{\sup_{\{i_1 \wedge \dots \wedge i_{k-1} \wedge i_{k+1} \wedge \dots \wedge i_q\}}}$$

Definition 9. Let λ be a pre-defined minimum confidence threshold. A rule R is called a highly coherent association rule (abbreviated as *HCAR*) if $conf_R \geq \lambda$.

Based on above definitions, in this study, the problem of mining highly coherent association rules can be defined as follows. Assume a database contains a number of transactions, and each transaction is recorded with the items purchased. Also, a contingency table is given to define the four kinds of support values for each possible pair in an itemset, as described in Definition 8. The problem is to first find itemsets with high support values larger than or equal to a predefined minimum support threshold, and then highly coherent itemsets could be induced from the set of frequent itemsets. Finally, the rules with high confidence values satisfying a predefined minimum confidence threshold can be found from the set of highly coherent itemsets.

To speed up the execution efficiency of finding highly coherent rules, the paper presents a projection-based coherent rule mining algorithm (named *PCMA*) to solve this problem.

4 Proposed Mining Algorithm

In this session, an effective recognizing strategy used in the proposed *PCMA* algorithm is described, and the strategy is first described in section 4.1. The proposed approach is then stated in section 4.2. A *Finding-FI* procedure of the proposed approach is given in section 4.3.

4.1 The Recognizing Strategy for Unpromising Itemsets

The recognizing strategy is developed to early determine whether or not an itemset is a highly coherent itemset. The main concept behind the strategy is that according to the proprieties of coherent rules defined in the previous section, for a coherent rule $X \rightarrow Y$, let the supports of X and Y be P_1 and P_2 , respectively, minimum confidence be $\lambda \geq 0.5$ and the support of transactions be 1. In the following, assume that the derived coherent rule $X \rightarrow Y$ has the highest confidence if $P_1 > P_2$. The contingency table is shown in Table 4.

Table 4. Contingency table for $P_1 > P_2$.

Frequency of co-occurrences		Consequence Y	
		Y	$\neg Y$
Antecedent X	X	P_2	$P_1 - P_2$
	$\neg X$	0	$1 - P_1$

From Table 4, according to the four criteria, if the rule $X \rightarrow Y$ is a coherent rule, the following two inequalities must be satisfied:

- (1) $P_2 > (P_1 - P_2)$; (2) $(1 - P_1) > (P_1 - P_2)$; (3) $P_2 / P_1 > \lambda$; (4) $\text{Sup}(X, Y) \leq P_2$.

If $P_2 > P_1$, the contingency table can be represented as shown in Table 5.

Table 5. Contingency table for $P_2 > P_1$

Frequency of co-occurrences		Consequence Y	
		Y	$\neg Y$
Antecedent X	X	P_1	0
	$\neg X$	$P_2 - P_1$	$1 - P_2$

From Table 5, the following two inequalities must be satisfied:

- (5) $P_1 > P_2 - P_1$; (6) $(1 - P_2) > (P_2 - P_1)$; (7) $P_1 / P_2 > \lambda$; (8) $\text{Sup}(X, Y) \leq P_1$.

From inequalities (1) to (8), the following theorem is obtained.

Theorem 1. Let $X \rightarrow Y$ be a coherent rule, where X and Y are itemsets. Assume that the support of X is P_1 , the support of Y is P_2 , and the total number of transactions is T . Then, the support of Y must be in the range of $\text{Max}[(2 * P_1 - I), \lambda * P_1] < P_2 < \text{Min}[(P_1 + I) / 2, (1 / \lambda) * P_1]$.

Proof: If $P_1 > P_2$, then (1) $P_2 > (P_1 - P_2)$, (2) $(1 - P_1) > (P_1 - P_2)$, (3) $P_2 / P_1 > \lambda$ and (4) $\text{Sup}(X, Y) \leq P_2$, which can be transformed into (1') $P_2 > 0.5 * P_1$, (2') $P_2 > (2 * P_1 - I)$, (3') $P_2 > \lambda * P_1$ and (4') $\text{Sup}(X, Y) \leq P_2$. If $P_1 < P_2$, then (5) $P_1 > P_2 - P_1$, (6) $(1 - P_2) > (P_2 - P_1)$, (7) $P_1 / P_2 > \lambda$ and (8) $\text{Sup}(X, Y) < P_1$, which can be transformed into (5') $2 * P_1 > P_2$, (6') $P_2 < (P_1 + I) / 2$, (7') $P_2 < (1 / \lambda) * P_1$ and (8') $\text{Sup}(X, Y) \leq P_1$. Thus, from (2') and (6'), (9) $(2 * P_1 - I) < P_2 < (P_1 + I) / 2$. From (3') and (7'), (10) $\lambda * P_1 < P_2 < (1 / \lambda) * P_1$. From (9) and (10), the lower and upper bounds of Y are derived as $\text{Max}[(2 * P_1 - I), \lambda * P_1]$ and $\text{Min}[(P_1 + I) / 2, (1 / \lambda) * P_1]$, respectively.

Thus, given an itemset S , its candidate coherent itemsets $\{X, Y\}$ can be formed, where X and Y are subsets of S , and $(X \cap Y) = \emptyset$. Calculating the contingency table of a candidate coherent itemset is time-consuming. To speed it up, by using Theorem 1, if the support of Y is not in the interval, then such itemset cannot generate coherent rules and it is removed directly (i.e., its contingency table is not calculated).

4.2 Proposed Projection-Based Coherent Rule Mining Algorithm (PCMA)

INPUT: A set of items, each with a value; a transaction database D , in which each transaction includes a subset of items; a minimum support threshold min_sup α ; a minimum confidence threshold min_conf λ .

OUTPUT: A final set of highly coherent association rules, $HCARs$.

STEP 1: For each item i in D , do the following substeps.

- (a) Scan the database and calculate the support of each item i .
- (b) Compare the support value of each item i , Sup_i , with to the predefined minimum support α . If the support value of item i is larger than or equals to the minimum support threshold, then put it into large l -itemset as follows:

$$L_l = \{i \mid \text{Sup}_i \geq \alpha, 1 \leq j \leq k\}.$$

STEP 2: For each y -th transaction Trans_y in D , do the following substeps.

- (a) Get each item i in Trans_y .
- (b) Check whether i appears in L_l or not. If it does, then keep the item i in the transaction Trans_y ; otherwise, remove the item i from Trans_y .
- (c) If the number of items kept in the modified transaction Trans_y is less than the value of 2, then remove the modified transaction Trans_y from D ; otherwise, keep it in TDB .

STEP 3: Process each item i in the set of L_l in alphabetical order of them by the following substeps.

- (a) Find the relevant transactions including i in D , and put the transactions in the set of projected transactions d_i of the item i .

- (b) Find all the frequent itemsets with i as their prefix item by the *Finding-FI*(i, d_i, I) procedure, and put returned frequent itemsets with the prefix i in L_{all} .

STEP 4: Initialize the temporary coherent itemset TCI table as an empty table, in which each tuple consists of two fields: itemset and the support of itemset.

STEP 5: Put all itemsets with more than one item and their corresponding supports in the set L_{all} into the temporary coherent itemset TCI table.

STEP 6: For each frequent I -itemset i in L_I , find the lower bound (LB_i) and upper bound (UB_i) of the support Sup_i by following formula:

$$LB_i < Sup_i < UB_i,$$

where LB_i and UB_i are calculated by $LB_i = \text{Max}[(2 * Sup_i - I), \lambda * Sup_i]$ and $UB_i = \text{Min}[(Sup_i + I) / 2, (1 / \lambda) * Sup_i]$. Note that the main aim of the two bounds (LB_i and UB_i) of i is to check whether an itemset S consisting of two subsets X and Y is a promising coherent itemset or not, where the first item in X is item i . That is, if the support of the subset Y in S is not within the interval $[LB_i, UB_i]$, then the S must impossible to be a highly coherent itemset.

STEP 7: For each itemset S in the TCI table, do the following substeps.

- (a) Generate all possible pairs of sub-itemsets X and Y from the itemset S , where X and Y are subsets of S , $(X \cup Y) = S$, and $(X \cap Y) = \emptyset$.
- (b) Check whether the Sup_Y for each possible pair (X, Y) is in the interval $[LB_i, UB_i]$ of the first item i in X . If the itemset S exists at least a pair (X, Y) , which the Sup_Y for the pair (X, Y) is not in the interval $[LB_i, UB_i]$ of the first item i in X , then remove the S from the TCI table; Otherwise, omit the S . Note that since the two subsets X and Y for a pair (X, Y) can be exchanged each other in the row and column fields of the contingency table; the results for the two pairs (X, Y) and (Y, X) in the contingency table are the same.

STEP 8: For each itemset S kept in the TCI table, do the following substeps.

- (a) Generate all possible pairs of sub-itemsets X and Y from the itemset S . Note that the superset of the two itemsets X and Y for any pair (X, Y) must be S .
- (b) For each pair (X, Y) , calculate four support values for antecedent X and consequent Y in the contingency table by using the set L_{all} , including Q_1 : Sup_{XY} , Q_2 : $Sup_{X \rightarrow Y}$, Q_3 : $Sup_{\neg XY}$ and Q_4 : $Sup_{\neg X \rightarrow Y}$. Note that Q_2, Q_3 and Q_4 can be calculated by $(Sup_X - Q_1)$, $(Sup_Y - Q_1)$, and $(I - Q_1 - Q_2 - Q_3)$, respectively.
- (c) If the itemset S exists a pair (X, Y) , which does not meet all the four conditions, which are $Q_1 > Q_2$, $Q_1 > Q_3$, $Q_4 > Q_2$, and $Q_4 > Q_3$, then S is omitted; Otherwise, itemset S is put into the set of highly coherent itemsets, HC .

STEP 9: For each itemset S in the set of HC , do the following substeps:

- (a) Generate its all candidate coherent association rules $(X \rightarrow Y)$, where X and Y are subsets of S , $(X \cup Y) = S$, and $(X \cap Y) = \emptyset$.

- (b) Calculate the confidence value of each candidate coherent association rule $(X \rightarrow Y)$ using $Conf_{X \rightarrow Y} = Sup_{XY} / Sup_X$.
- (c) Check each rule $(X \rightarrow Y)$ whether its confidence value $Conf_{X \rightarrow Y}$ is larger than or equals to the minimum confidence threshold λ or not. If yes, put the rule $(X \rightarrow Y)$ into the set of highly coherent association rules (HCARs); otherwise, omit the rule.

STEP 10: Output the highly coherent association rule set HCARs.

After STEP 3, all frequent itemsets in the database are found. The *Finding-FI*(X, d_X, r) procedure finds all the frequent itemsets with the r -itemset X as their prefix itemsets, and the procedure is stated below.

4.3 The Finding-FI(X, d_X, r) Procedure

Input: A prefix r -itemset X and its corresponding projected transactions d_X .

Output: The frequent itemsets with X as their prefix itemsets.

STEP 1: Initialize the temporary itemset TI_X table as an empty table, in which each tuple consists of two fields: itemset and the actual support of the itemset.

STEP 2: For each y -th transaction $Trans_y$ in d_X , do the following substeps.

- (a) Get each item i located after X in $Trans_y$.
- (b) Generate the $(r+1)$ -itemset S composed of the prefix r -itemset X and I , put it in the TI_X table, and add the value of 1 to its count field value in the TI_X table.

STEP 3: For each $(r+1)$ -itemset S in the TI_X table, check support value of S , Sup_S , against to the predefined minimum support threshold α . If the support value of S is larger than or equals to the minimum support threshold, then put it into the set $L_{(r+1), X}$ of large $(r+1)$ -itemsets with the prefix X .

STEP 4: Acquire the items appearing in the set of $L_{(r+1), X}$ of X , and put them in the set of promising items, $PI_{(r+1), X}$.

STEP 5: Set $r = r + 1$, where r represents the number of items in the processed subitemsets.

STEP 6: For each y -th transaction $Trans_y$ in d_X , do the following substeps.

- (a) Check whether each item I in $Trans_y$ appears in $PI_{r, X}$ or not. If it does, then keep the item i in $Trans_y$; remove the item i from $Trans_y$.
- (b) If the number of items kept in the modified transaction $Trans_y$ is less than the value $(= r + 1)$, remove the modified transaction $Trans_y$ from d_X ; otherwise, kept it in d_X .

STEP 7: Process each itemset S in the set of $L_{(r+1), X}$ of X in alphabetical order of them by the following substeps.

- (a) Find the relevant transactions including S from d_X , and then put the transactions including S in the set of projected transactions d_S of prefix S .
- (b) Find all frequent itemsets with S as their prefix itemset by the *Finding-FI*($S, d_S, r + 1$) procedure. Let the set of returned frequent itemsets with X as their prefix itemsets be L_X .

STEP 8: Return the frequent itemsets in the set of FI_X .

5 Experimental Results

This section presents the experimental results of the proposed approaches. The programs were implemented in Java on a personal computer with an Intel Core2 2.2-GHz CPU and 2 GB of RAM running Windows 7 Professional. These algorithms used a real data set, the foodmart data set, which was stored on a Microsoft SQL Server. It contains 21,557 transactions and 1,559 items. In the experiments, the minimum support threshold was varied to study its effect.

The foodmart data set was used to evaluate the efficiency and performance of the proposed approach. Two minimum supports were used to compare the derived rules obtained using the proposed approach (PCMA), *HCARM* and the Apriori algorithm. The results are shown in Table 6. The minimum supports were set at 0.01% and 0.015%.

Table 6. Comparison between PCMA and other approaches

Approach	Minimum support	Total number of rules	Average conf.	Execution time (sec.)
<i>PCMA</i>	0.01%	2,840	0.918	2.888
	0.015%	136	0.906	2.816
<i>HCARM</i>	0.01%	2,840	0.918	2,739
	0.015%	136	0.906	2,481
<i>Apriori</i>	0.01%	18,880	0.325	2,783
	0.015%	2,520	0.211	2,440

Table 6 shows that the numbers of rules derived by the PCMA and *HCARM* algorithms are both 2840 and 136 for supports of 0.01% and 0.015%, respectively. Although fewer rules are derived compared to those obtained using the Apriori approach, these results are reasonable because the PCMA algorithm has stricter constraints. In addition, the average confidence levels of the derived rules obtained using PCMA are 0.918 and 0.906 and those obtained using the Apriori approach are 0.325 and 0.211, respectively. The table shows that the execution time of PCMA is lower than the other two algorithms and does not increase dramatically with decreasing of minimum supports. These results show that the rules derived by PCMA are more interesting to business due to their high confidence values. And, the execution time of the proposed approach is efficient than that by *HCARM*.

6 Conclusion and Future Works

In this paper, we have proposed algorithm for mining highly coherent rule from transaction database. The algorithm is named projection-based coherent rule mining algorithm (PCMA), which is utilizing projection technique for improving the efficiency of the mining process and further speed up the execution time of processing than apriori-based algorithm. In the approach, the developed algorithm can provide

more interesting and reliable rules to market decision makers for making marketing strategies accurately. Since generating candidate coherent itemsets is time-consuming, the lower and upper bounds of the itemsets are deduced for removing itemsets that cannot generate highly coherent rule. Without generating extra candidate itemsets, it can reduce times in searching database and thus save a lot of execution times. Experimental results on a real dataset are made to show the validation performance of the proposed approaches. In the future work, we will focus on extending the proposed approach to more complex problems.

Acknowledgement. This research was supported by the National Science Council of the Republic of China under contract NSC 102-2923-E-390-001-MY3.

References

1. Agarwal, R., Aggarwal, C., Prasad, V.V.V.: A Tree Projection Algorithm for Generation of Frequent Itemsets. *Journal of Parallel and Distributed Computing* 61(3), 350–371 (2001)
2. Agrawal, R., Imielinski, T., Swami, A.: Mining Association Rules between Sets of Items in Large Databases. In: *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 207–216 (1993)
3. Bie, T.D.: An Information Theoretic Framework for Data Mining. In: *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 564–572 (2011)
4. Brin, S., Motwani, R., Silverstein, C.: Beyond Market Baskets: Generalizing Association Rules to Correlations. In: *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 265–276 (1997)
5. Cheung, Y.L., Fu, A.W.C.: Mining Frequent Itemsets without Support Threshold: With and Without Item Constraints. *IEEE Transactions on Knowledge and Data Engineering* 16(9), 1052–1069 (2004)
6. Chen, C.H., Lan, G.C., Lin, Y.K., Hong, T.P.: Mining high coherent association rules with consideration of support measure. *Expert Systems with Applications* 40(16), 6531–6537 (2013)
7. Chiang, D.A., Wang, C.T., Chen, S.P., Chen, C.C.: The Cyclic Model Analysis on Sequential Patterns. *IEEE Transactions on Knowledge and Data Engineering* 21(11), 1617–1628 (2009)
8. Plantevit, M., Laurent, A., Laurent, D., Teisseire, M., Choong, Y.W.: Mining Multidimensional and Multilevel Sequential Patterns. *ACM Transactions on Knowledge Discovery from Data* 4(1), 4–37 (2010)
9. Sim, A.T.H., Indrawan, M., Zutshi, S., Srinivasan, B.: Logic-Based Pattern Discovery. *IEEE Transactions on Knowledge and Data Engineering* 22(6), 798–811 (2010)
10. de Sá, C.R., Soares, C., Jorge, A.M., Azevedo, P., Costa, J.: Mining Association Rules for Label Ranking. In: Huang, J.Z., Cao, L., Srivastava, J. (eds.) *PAKDD 2011, Part II*. LNCS, vol. 6635, pp. 432–443. Springer, Heidelberg (2011)
11. Wang, K., He, Y., Han, J.: Pushing Support Constraints into Association Rules Mining. *IEEE Transactions on Knowledge and Data Engineering* 15(3), 642–658 (2003)

ICISLM: Design of an Integrated Cloud Information System for Logistic Management Based on Web Server Virtualization

Shang-Liang Chen¹, Yun-Yao Chen^{1,*}, Hsuan-Pei Wang¹, and Chiang Hsu²

¹ Institute of Manufacturing Information and Systems, National Cheng Kung University, Tainan City, Taiwan, R.O.C.

² Department of Business Administration, Chang Jung Christian University, Tainan City, Taiwan, R.O.C.

slchen@mail.ncku.edu.tw, yewyewchen@gmail.com,
heart.1416@hotmail.com, chiangh@mail.cjcu.edu.tw

Abstract. This research aims to design an integrated cloud information system for logistic management (ICISLM) based on web server virtualization and Software as a Service (SaaS) architecture. Based on web server virtualization technology, ICISLM provides rentable and scalable environments for enterprises, enabling enterprise customers to operate the system without the need to install any software or manage the cloud servers. The aim of ICISLM is to provide a single system that can satisfy warehouse management, logistics monitoring, and downstream customer needs. A workflow dispatch mechanism was designed to dynamically dispatch the services host and to stabilize system operation. A case study is provided to verify the function of logistics service pools.

Keywords: web server virtualization, cloud computing, SaaS, RFID, ICISLM.

1 Introduction

The rapid development of large-scale server hosts and improvements in their processing capacity in recent years have allowed the replacement of a large number of individual computers. Properly utilized, virtualization technology can be applied to improve the performance of large-scale servers. The pervasiveness of Web 2.0 and Wi-Fi networks has also quickened facilitated service integration and resource sharing around the world. In previous study, Chen et al. proposed a SaaS model for implementing cloud-based logistic systems, which enabled enterprises to develop its own logistic systems in both private and public cloud infrastructures [1]. Based on the their research, this study aimed to provide logistics enterprises with an integrated cloud information system for logistics management by utilizing SaaS architecture as

* Corresponding author.

the cloud service provider with virtual server technology. The system deployed in this study uses virtualization technology to build a single server host with multiple virtual web server environments. This reduces the cost of hardware deployment and improves operational efficiency.

1.1 Application in Cloud Computing for Logistic Information Systems

The NIST (National Institute of Standards and Technology) indicates that a cloud computing system must include five basic characteristics: (1) On-demand self-service, (2) broad network access, (3) resource pooling, (4) rapid elasticity, and (5) Measured Service [2]. We discuss the present cloud service company to include VMWare [3], Chunghwa Telecom Hicloud (CAAS) [4] and Microsoft MCloud (OACloud cloud Office) [5] and a method by which to achieve the five basic characteristics suggested by the NIST. We have found that many cloud service companies are providing similar services, but few provide for logistics management. Bernhard Holtkamp, etc. analyzed the three service models defined by the NIST for the cloud logistics management system [6], and they suggested that the system should consider more than one warehouse worker to share a server, for example, implementing a user identity mechanism to avoid termination of the application. Shang-Liang Chen etc. proposed a model based on SaaS cloud computing architecture for the RFID logistics management system [7]. They adopted dual-mode RFID to develop logistics and inventory systems and used service-oriented architecture to enhance system flexibility [8]. However, according to the SaaS maturity model levels proposed by the Microsoft MSDN Architecture Center [9], we established that the above research only matched the SaaS maturity model to level-2. Thus, it is the goal of the current work to develop a rentable cloud system for logistics management and to improve the architecture in order to satisfy the SaaS maturity model to level-4. According to the definition from Forrester's SaaS Maturity Model level 4 [10]: *At level 4, an advanced SaaS vendor provides not only a well-defined business application but also a platform for additional business logic. This complements the original single application of the previous level with third-party packaged SaaS solutions and even custom extensions. The model even satisfies the requirements of large enterprises, which can migrate a complete business domain like "customer care" toward SaaS.*

1.2 Application of Virtual Web Server Technology

Microsoft has pointed out that most servers have an average utilization rate of only 10-15%. Hence, virtualizing the server can save more than 60 percent of the hardware costs [11]. A comparison of the virtual and physical web server architecture [12] [13] is provided in Table 1:

Table 1. Compare with virtual web server architecture and the physical web server architecture

Item	Physical Server	Web	Virtual Web Server
Deployed	Single deployment entry and poor efficiency.		A physical machine can support multiple operating systems and web server environments; decentralized deployment to enhance efficiency and utilization.
Flexibility	(1) Low flexibility and scalability. (2) Hard to create the same operating environment quickly.		(1) Higher flexibility and expandability. (2) It's easy to create the same operating environment quickly. (3) Flexible allocation of computer resources such as hard disk space and memory. (4) Easy transition between the different entities of the server.
Computing	Higher computing performance may be attained, but the CPU and memory may be underused.		Appropriate allocation of CPU and memory may improve the utilization of a single server.

2 Research Methods

2.1 Cloud Service Role of ICISL

The ICISLM system architecture is shown in Figure 1, indicating the relationship between the renting enterprises and access to ICISLM and is described as follows:

1. Cloud Provider Domain

Cloud Provider Domain refers to the cloud infrastructure and its logistics services. It is also regarded as the core focus of this study and consists of four items:

- (a) Logistics Service Pools (LSP): An LSP is deployed on the virtual web server (VWS). Logistics application services are deployed on a VWS and are encapsulated with LSPs to provide enterprise access through a customized ICISLM interface.
- (b) Virtual Web Server for Enterprises (VWSE): VWSE refers to the virtual web server that businesses can rent. Services are quantified according to lease duration, CPU number, RAM size and the usage of net traffic. Therefore, the VWSE deploys a customized ICISLM interface, an authentication mechanism and enterprise-self-deployment applications.

- (c) Data Center (DC): A DC is adopted for centralized service and infrastructure management. The virtual machine server hardware is described in Section 3.
- (d) Backup Server (BS): A BS is constructed to store backup files from the LSP, VWSE and DC resources.

2. Enterprise Domain

The services or applications deployed by the enterprises themselves are called global services.

3. User Domain

This refers to the user access to the service through the application. In this study, users can access the logistic services on LSPs with all web browsers (IE, Firefox and Chrome). The LSP and VWSE are both deployed on multiple virtual web servers, and the amount of VWSE is decided by the number of enterprises renting the service. VWSEs are distributed with a back-end operation mode by a cloud provider. In this study, the system architecture satisfies three characteristics of the SaaS maturity model level-4 standard, including rentability, centralized management and dynamic allocation of resources based on a workload dispatch mechanism described in the next section.

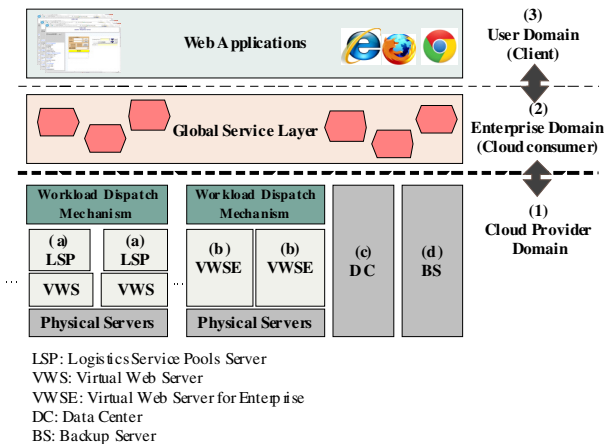


Fig. 1. ICISLM Architecture

The design of the workload dispatch mechanism (WDM) is shown in Figure 2. The WDM determines whether there is a system overload when a user requests access to ICISLM (User Authentication) and logistics services (LSP Request). The overload computing module is asynchronously operated in background mode to smooth the system interface operation. In addition, a Capture vNIC Traffic Service (CVNTS) was designed to capture the virtual web server net traffic as well as to write to the database

(VWSE Group/LSP Group Database). CVNTS is a continuously running background program designed with an overload computing algorithm in this paper, the pseudo code of overload computing algorithm is shown in Table 2. The algorithm provides CVNTS with the function to extract data from the overload computing module. If the OverLoad variable is true, this system automatically redirects to the next virtual web server. However, if the next virtual web server does not exist, the ICISLM system stays in the original server and sends an SMS to the customer or cloud provider.

Table 2. Pseudo code of overload computing algorithm

Denote the *OverLoad* initial value is **false**
IF ($TrafficCurrent > UpperTraffic$ AND $UserOnline > UserUpper$)
 Set *OverLoad* as **True**;
OverLoad: Record the status of overload, the variable type is Boolean.
TrafficCurrent: Record of the value of current net traffic.
UserOnline: Record of the value of current online user
TrafficUpper: The upper threshold of net traffic
UserUpper: The upper threshold of the number of online user

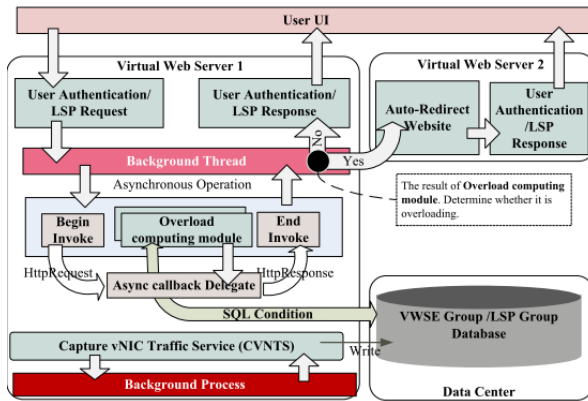


Fig. 2. Design of WDM

2.2 Design of Logistics Services Pool (LSP)

In this study, we organized the complex information among logistics roles into a set of service processes so that users can easily integrate all of the information in the identity permissions. The application scenario contains three tiers: a cloud provider tier (i.e. mark of A to D), an enterprise role tier (i.e. mark of E to G) and an application process scenario tier (i.e. labelled 1 to 5). However, the application scenario explanations are as follows:

- 1. Inventory and Stocking:** Suppliers use RFID technology to save information about goods and inventory. At first, we use the UHF RFID for the pallets inventory and then use the HF RFID for goods included in the pallet inventory. The inventory results will save to the system database.

2. **Shipping:** Shipping is a relationship between the supplier and the logistics company. Pallets will be automatically inventoried again with the UHF RFID deployed at the exit and will confirm the content of the order form.
3. **Logistic tracking:** Tracks the delivery path and save the GPS/GIS information to the database.
4. **Electronic receipt:** A relationship between the logistic company and customer. Customer use a handheld device to sign their names, thus replacing paper and pen.
5. **Search:** Customers can use system to search where their product is. This function is provided by the Google maps API.

Therefore, in this study, we designed five logistics services and then combined them into a logistics service pool (LSP):

1. **User Management Services Pool:** The main function of this service pool is the management of user permissions and the personal information of users.
2. **Message Pushing Service Pool:** The main function of this service pool is to implement short message service (SMS) via the rented Dynamic-Link Library (DLL) service from a Telecommunication Company.
3. **Supply Management Services Pool:** The main function of this service pool is the maintenance of cargo information and the shipped inventory. In this phase, we use the RFID device to support our system. The HF RFID is used to inventory goods, and the UHF RFID is used inventory pallets. Initially, the staff inventories pallets via the UHF RFID and then saves the pallet information to the database. A virtual com port technology was proposed to provide a connection between the RFID hardware and the database [12]. In the second step, the staff prepares the goods according to the order content, and then inventories these goods via HF RFID. When the stocking is completed, these pallets are sent to the conveyor belt for the second confirmation; then, through the UHF RFID tag on the pallet, are sent to inventory again.
4. **Logistics Monitor Service Pool:** The current location of the goods is automatically compared to its final destination and displayed on Google Maps.
5. **Net Traffic Monitor Service Pool:** This service allows VWSE customers to monitor network traffic. In addition, the value of net traffic will be stored in the database automatically every second so that further results can be computed by the WDM.

2.3 Design of ICISLM

This section focuses on ICISLM functions and user interface design for the purpose of verification. Two sub systems is designed for a case study for ICISLM. Users can rent logistic services with the following interfaces based on a implementation of ICISLM.

The shipped inventory subsystem:

Figure 3 shows the pallet inventory interface. The supplier uses this service to prepare the pallet, automatically inventoried by the UHF RFID reader, for shipping.

Goods are prepared and added to the pallet in the next step, where electronic tags (specification: ISO 14443A) are pasted on, and an HF RFID reader is used to inventory them.

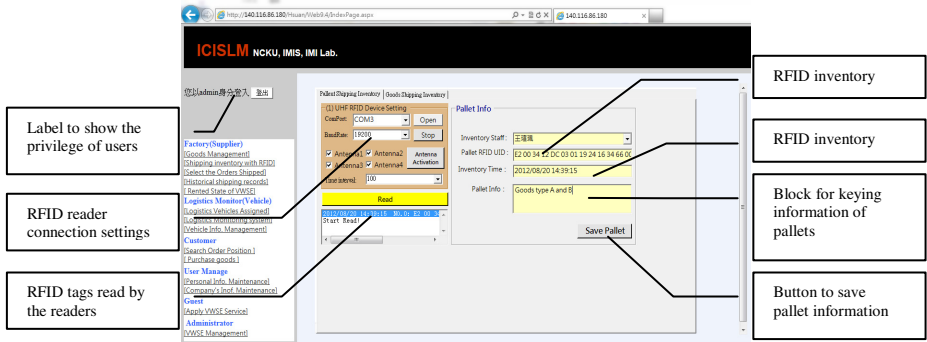


Fig. 3. ICISLM user interface for pallet inventory

The logistics management and monitoring subsystem:

The logistics monitoring interface is shown in Figure 4. This subsystem shows the work of the logistics staff and the goods position on the map based on GPS signals. In addition, a function to prompt the logistics vehicle automatically if it is near the destination is provided. Then, a receipt is sent to the customer, and an SMS message is then sent to the customer upon signing the receipt.

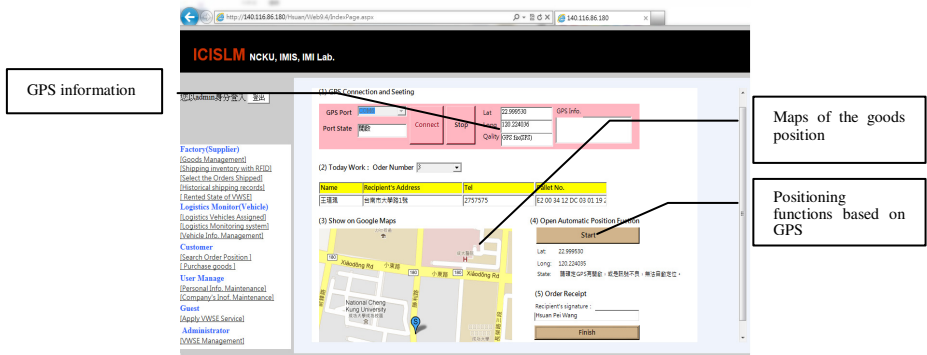


Fig. 4. ICISLM user interface for logistic monitoring

3 Conclusions

Important achievements and contribution of this study including (1) design of a rentable virtual web server environment. Provision of an independent ICISLM for customers to rent that allows easy deployment of web applications. (2) design of

logistic service pools for logistics business characterized by flexibility and adaptivity to customer needs. (3) design of a workload dispatch mechanism for the virtual web servers for the provision of a stable virtual web server environment for business customers to rent. (4) design of an ICISLM system which satisfies SaaS model, which is rentable and scalable environments for enterprises, enabling customers to operate the system without the need to install any software or manage the cloud servers. Implementation and experiments will be done in our future works.

Acknowledgments. This paper was supported by the Ministry of Science and Technology (MOST 102-2221-E-006-107). This support was crucial to the successful completion of this research project.

References

1. Chen, S.L., Chen, Y.Y., Hsu, C.: A New Approach to Integrate Internet-of-Things and Software-as-a-Service Model for Logistic Systems: A Case Study. *Sensors* 2014, 6144–6164 (2014)
2. Mell, P., Grance, T.: A NIST Definition of Cloud Computing. National Institute of Standards (2011)
3. VMWare, <http://www.vmware.com/tw>
4. hicloud (CAAS), <http://hicloud.hinet.net/index.html>
5. 7MCloud (Microsoft), <http://www.microsoft.com/taiwan/mcloud/>
6. Holtkamp, B., Steinbuss, S., Gsell, H., Loeffeler, T., Springer, U.: Towards a Logistics Cloud. In: 2010 Sixth International Conference on Semantics, Knowledge and Grids, pp. 305–309 (2010)
7. Chen, S.L., Chen, Y.Y.: Design and Implementation of a Global Logistic Tracking System Based on SaaS Cloud Computing Infrastructure. *Journal of System and Management Sciences* 1, 85–96 (2011)
8. Chen, S.L., Chen, Y.Y., Hsu, C.: Development of Logistic Management Information System Based on Web Service Architecture and RFID Technology. *Applied Mathematics & Information Sciences* 7, 939–946 (2013)
9. Carraro, G., Chong, F.: Software as a Service (SaaS): An Enterprise Perspective. Microsoft MSDN Architecture Center (2006)
10. Stefan, R.: Forrester's SaaS Maturity Model, <http://www.forrester.com/Forresters+SaaS+Maturity+Model/fulltext/-/E-RES46817>
11. Microsoft: Server Virtualization, <http://www.microsoft.com/taiwan/savemoney/vsv1.htm>
12. Hoffman, J.: Virtualization: Virtualization in and Beyond the Cloud. *TechNet Magazine* (2012)
13. Stagner, H.: Web server virtualization done right, TechTarget Sites, Topics: Virtualization Technology and Services Server (2007)
14. Chen, S.L., Wang, H.P., Chen, Y.Y., Hsu, C.: Development of Software-as-a-Service Cloud Computing Architecture for Manufacturing Management Systems Based on Virtual COM Port Driver Technology. In: *Applied Mechanics and Materials*, vol. 479, pp. 1023–1026 (2013)

Hiding Sensitive Itemsets with Minimal Side Effects in Privacy Preserving Data Mining

Chun-Wei Lin^{1,2}, Tzung-Pei Hong^{3,4,*}, and Hung-Chuan Hsu³

¹Innovative Information Industry Research Center (IIIRC),

²Shenzhen Key Laboratory of Internet Information Collaboration

School of Computer Science and Technology

Harbin Institute of Technology Shenzhen Graduate School

HIT Campus Shenzhen University Town, Xili, Shenzhen 518055 P.R. China

³Department of Computer Science and Information Engineering

National University of Kaohsiung, Kaohsiung, Taiwan, R.O.C.

⁴Department of Computer Science and Engineering

National Sun Yat-sen University, Kaohsiung, Taiwan, R.O.C.

jerrylin@ieee.org, tphong@nuk.edu.tw, crsbird04@gmail.com

Abstract. Privacy-preserving data mining (PPDM) has become an important issue to hide the confidential or private data before it is shared or published in recent years. In this paper, a novel algorithm is proposed to hide sensitive itemsets through item deletion. Three side effects of hiding failures, missing itemsets, and artificial itemsets are considered to evaluate whether the transactions or the itemsets are required to be deleted for hiding sensitive itemsets. Experiments are then conducted to show the performance of the proposed algorithm in execution time, number of deleted transactions, and number of side effects.

Keywords: Privacy-preserving data mining, side effects, information hiding, data sanitization, sensitive itemsets.

1 Introduction

With the rapid growth of data mining technologies in recent years, useful information can be easily mined out to aid managers or decision-makers [1-2, 6, 13, 16, 20]. In some applications, the confidential or secure data is required to be hidden before it is shared or published. Privacy-preserving data mining (PPDM) [4] was thus proposed to reduce privacy threats by hiding sensitive information while allowing required information to be mined out from databases. The intuitive way for hiding sensitive information of data sanitization in PPDM is directly to delete sensitive information from amounts of data. Amiri thus proposed the aggregate approach, disaggregate approach, and hybrid approach to determine whether items or transactions are required to be deleted for hiding sensitive information [5]. The infrequent itemset is, however, not considered in the evaluation process, thus raising the probability of

* Corresponding author.

artificial itemsets caused. Besides, the differences between the minimum support threshold and the frequencies of the itemsets to be hidden are not considered in the above approaches.

In this paper, a hiding-missing utility (HMU) algorithm is thus proposed to hide the user-specified sensitive itemsets through item deletion. The proposed HMU algorithm is to delete the items for hiding the sensitive itemsets, thus avoiding to change the number of transactions in the sanitized databases. Only the hiding failures and missing itemsets are concerned to evaluate the transactions for deleting the items but the artificial itemsets. Experiments are then conducted to show the performance of the proposed algorithm in execution time, number of deleted transactions, and number of side effects, compared to the disaggregate approach [5].

2 Review of Related Works

In this section, the data mining techniques and privacy preserving data mining techniques are respectively reviewed.

2.1 Data Mining Techniques

Data mining is used to extract useful rules from large amounts of data [1-2, 6, 16, 18, 20] and the most commonly one is to derive association rules [1-2]. In association-rule mining, it consists of two phases for firstly finding the frequent itemsets by a generate-and-test approach then secondly generating the association rules by the combination approach [2]. Instead of mining single-level association rules, Han and Fu proposed ML_TILA algorithm to discover multiple-level association rules [11]. Han et al. proposed the Frequent-Pattern-tree (FP-tree) structure for efficiently mining association rules without generation of candidate itemsets [12]. The FP-tree was used to compress a database into a tree structure which stored only large items. It was condensed and complete for finding all the frequent patterns. They showed the FP-tree approach could have a better performance than Apriori. Other approaches to efficiently mine the desired information are still developed in progress.

2.2 Privacy Preserving Data Mining

Information can be efficiently discovered using various data mining techniques. The misuse of these techniques, however, may lead to privacy concerns and security problems. Privacy-preserving data mining (PPDM) has thus become a critical issue for hiding private, confidential, or secure information [3, 19, 22]. In PPDM, data sanitization is an intuitive way to directly delete sensitive data for hiding sensitive information. Leary found that data mining techniques can pose security and privacy threats [17]. Amiri thus proposed the aggregate, disaggregate, and hybrid approaches to respectively determine the whether the transactions or the items to be deleted for hiding sensitive information [5]. Divanis and Verykios protected sensitive information by hiding sensitive itemsets instead of hiding sensitive rules [8].

Modi et al. proposed a Decrease Support of R.H.S. item of Rule Clusters (DSRRC) algorithm to hide all sensitive rules by deleting R.H.S items of the sensitive rules [21]. Some studies are designed to add noise information [9-10], or modify transactions [7, 22] for the purpose of sanitization. Various PPDM strategies are still designed in progress to hide the sensitive information by different evaluation criteria [14-15, 19].

The relationships of itemsets before and after the sanitization process can be described as shown in Figure 1, where H represents the frequent itemsets in the original database, S represents the sensitive itemsets defined by users that are frequent itemsets but need to be protected, $\sim S$ represents the non-sensitive itemsets that are frequent itemsets, and H' represents the frequent itemsets after the sanitization process.

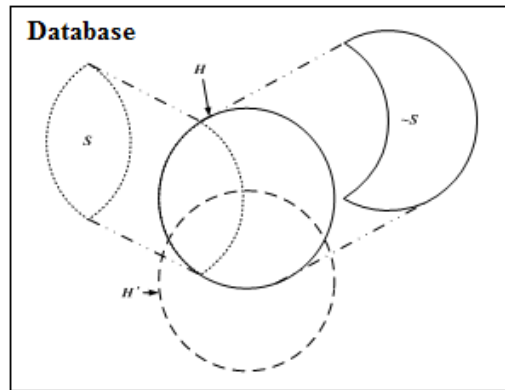


Fig. 1. Relationship between the side effects and mined rules

Let α be the number of sensitive itemsets that fail to be hidden after the sanitization process. Thus, α is the intersection of S and H' ($= S \cap H'$). The number of missing frequent itemsets is denoted as β . Thus, β is the difference between $\sim S$ and H' ($= \sim S - H'$). The number of artificial itemsets is denoted as γ . Thus, γ is the difference between H' and H ($= H' - H$).

3 Formula Definition

In the proposed formulas, the differences between minimum support threshold and the frequencies of the sensitive itemsets are thus considered to evaluate whether the transactions are required to be deleted instead of only the presences of the itemsets in the transactions. The details of two formulas are described below.

3.1 Hiding Failure Dimension

Hiding Failure Dimension (HFD) is used to evaluate the hiding failures of each processed transaction in the sanitization process. When a processed transaction T_k

contains a sensitive itemset hs_x , the HFD value of the processed transaction is calculated as:

$$HFD^k(hs_x) = \frac{MAX_{HS} - freq(hs_x) + 1}{MAX_{HS} - \lceil |D| \times \lambda \rceil + 1} \quad (1)$$

where λ is defined as the percentage of the minimum support threshold, sensitive itemset hs_x is from the set of sensitive itemsets HS , MAX_{HS} is the maximal count of the sensitive itemsets in the set of sensitive itemsets HS , $|D|$ is the number of transactions in the original database D , and $freq(hs_x)$ is the occurrence frequency of the sensitive itemset hs_x .

3.2 Missing Itemset Dimension

Missing Itemset Dimension (MID) is used to evaluate the itemsets of each processed transaction in the sanitization process. When a processed transaction T_k contains a frequent itemset fi_x , the MID value of the processed transaction is calculated as:

$$MID^k(fi_x) = \frac{MAX_{FI} - freq(fi_x) + 1}{MAX_{FI} - \lceil |D| \times \lambda \rceil + 1} \quad (2)$$

where an itemset fi_x is a frequent itemset from the set of large (frequent) itemsets FI , MAX_{FI} is the maximal count of the large itemsets in the set of FI , and $freq(fi_x)$ is the occurrence frequency of the large itemset fi_x .

4 Proposed HMU Algorithm through Item Deletion

The proposed HMU algorithm is described as follows.

Proposed HMU algorithm:

INPUT: An original database D , a minimum support threshold ratio λ , a risky bound μ , a set of large (frequent) itemsets $FI = \{fi_1, fi_2, \dots, fi_p\}$, and a set of sensitive itemsets to be hidden $HS = \{hs_1, hs_2, \dots, hs_r\}$.

OUTPUT: A sanitized database D^* with no sensitive information.

SETP 1: Select the transactions to form a projected database D' , where each transaction T_k in D' consists of sensitive itemsets hs_i within it, where $1 \leq i \leq r$.

STEP 2: Select the transactions without any of the sensitive itemsets to firstly form the sanitized database D^* .

STEP 3: Process each frequent itemset fi_j in the set of FI to determine whether its frequency satisfies the condition $freq(fi_j) = \lceil \lceil |D| \times \lambda \rceil \times (1 + \mu) \rceil$, where $|D|$ is the number of transactions in the original database D and $freq(fi_j)$ is the occurrence frequency of the large itemset fi_j . Put the fi_j that do not satisfy the condition into the set of FI_{imp} .

STEP 4: Calculate the *maximal count* (MAX_{HS}) of the sensitive itemsets hs_i in the set of HS as:

$$MAX_{HS} = \max\{freq(hs_i), \forall hs_i, 1 \leq i \leq r\},$$

where $freq(hs_i)$ is the occurrence frequency of the sensitive itemset hs_i in the set of HS .

STEP 5: Calculate the HFD of each transaction T_k . Do the following substeps:

Substep 5-1: Calculate the HFD of each sensitive itemsets hs_i within T_k as:

$$HFD^k(hs_i) = \frac{MAX_{HS} - freq(hs_i) + 1}{MAX_{HS} - \lceil |D| \times \lambda \rceil + 1}.$$

Substep 5-2: Sum the HFDs of sensitive itemsets hs_i within T_k as:

$$HFD^k = \frac{1}{\sum_{i=1}^r HFD^k(hs_i) + 1}.$$

Substep 5-3: Normalize the HFD^k for all transactions T_k in D' .

STEP 6: Calculate the maximal count (MAX_{FI}) of the large itemsets fi_j in the set of FI as:

$$MAX_{FI} = \max\{freq(fi_j), \forall fi_j, 1 \leq j \leq p\}.$$

STEP 7: Calculate the MID of each transaction T_k . Do the following substeps:

Substep 7-1: Calculate the MID of each large itemsets within T_k as:

$$MID^k(fi_j) = \frac{MAX_{FI} - freq(fi_j) + 1}{MAX_{FI} - \lceil |D| \times \lambda \rceil + 1}.$$

Substep 7-2: Sum the MIDs of large itemsets fi_j within T_k as:

$$MID^k = \sum_{j=1}^p MID^k(fi_j).$$

Substep 7-3: Normalize the MID^k for all transactions T_k in D' .

STEP 8: Calculate the HMAU for HFD and MID of each transaction T_k as:

$$HMU^k = w_1 \times HFD^k + w_2 \times MID^k,$$

where w_1 and w_2 are the pre-defined weights by users.

STEP 9: Calculate the number of 1-itemsets within the sensitive itemsets in HS as $nc(i_a)$.

STEP 10: Remove the maximal $nc(i_a)$ from T_k with the minimal HMU^k value as $\min\{HMU^k, \forall T_k, 1 \leq k \leq |D'|\}$ in D' , where 1-itemset i_b is appeared in T_k .

STEP 11: Update the occurrence frequencies of all sensitive itemsets in the set of HS .

STEP 12: Remove the hs_i in the set of HS if $freq(hs_i) < \text{minimum count}$ ($= \lceil |D| \times \lambda \rceil$).

STEP 13: Update the occurrence frequencies of all large itemsets in the set of FI and FI_{imp} .

STEP 14: Remove the fi_j if $freq(fi_j) < \text{minimum count}$ ($= \lceil |D| \times \lambda \rceil$), and put the fi_j into the set of FI .

STEP 15: Set the transaction without any of sensitive itemsets as a sanitized one. Put it from D' to D^* .

STEP 16: Repeat STEP 3 to STEP 15 until the set of HS is empty ($|HS| = 0$).

5 An Illustrated Example

Consider the database with 10 transactions (tuples) with 6 items (denoted as a to f) shown in Table 1. The minimum support threshold is initially set at 40%, and the risky bound is set at 10%. A set of sensitive itemsets, $HS = \{be:6, abe:4\}$, is considered to be hidden by the sanitization process.

Table 1. Original database

TID	Item	TID	Item
T1	a, b, c, e	T6	b, c, e
T2	e	T7	a, b, c, d, e
T3	b, c, e, f	T8	a, b, e
T4	d, f	T9	c, e
T5	a, b, d	T10	a, b, c, e

Based on an Apriori-like approach [2], the large (frequent) itemsets are $\{a:5, b:7, c:6, e:8, ab:5, ae:4, bc:5, be:6, ce:6, abe:4, bce:5\}$. After processing the designed algorithm, three dimensions of each transaction are shown in Table 2.

Table 2. Two dimensions of each transaction in projected database

TID	HFD	MRD	HMU
T1	0.57	1	0.785
T3	1	0.33	0.667
T6	1	0.33	0.667
T7	0.57	1	0.785
T8	0.57	0.67	0.619
T10	0.57	1	0.785

The designed algorithm is then performed to sanitize the original atabase. After all STEPs are processed, the final sanitized database D^* is obtained as shown in Table 3.

Table 3. Final sanitized database D^*

TID	Item	TID	Item
T1	a, b, c, e	T6	$b, c,$
T2	e	T7	a, b, c, d, e
T3	c, e, f	T8	a, b
T4	d, f	T9	c, e
T5	a, b, d	T10	a, b, c, e

6 Experimental Results

Experiments are conducted to show the performance of the proposed HMU algorithm compared to that of the disaggregate algorithm [5] for hiding sensitive itemsets

through item deletion. The real dataset BMS-WebView-1 [23] is used in the experiments. The minimum support thresholds are respectively set at 1% and 2%, and the percentages of sensitive itemsets are sequentially set at 5% to 25%. The risky bound in the experiments is initially set at 10%. Figure 2 shows the execution time for proposed HMU algorithm and disaggregate algorithm [5] at various sensitivity percentages of the frequent itemsets.

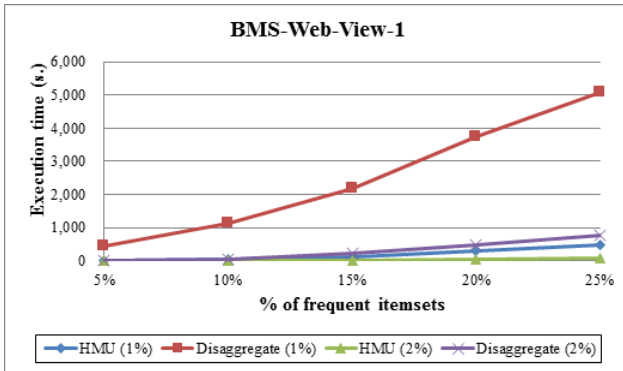


Fig. 2. Comparison of execution time of two different minimum support thresholds

From Figure 2, the proposed HMU algorithm has better performance the disaggregate algorithm. Experiments are also conducted to evaluate the number of deleted items. The results are shown in Figure 3.

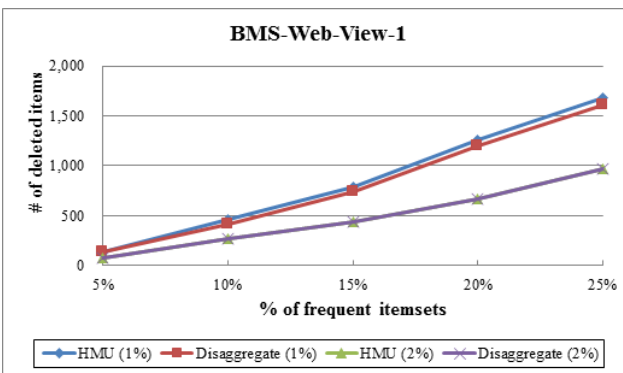


Fig. 3. Comparison of number of deleted items of two different minimum support thresholds

When the minimum support threshold is set at 1%, the proposed HMU algorithm is required to delete more items than the disaggregate algorithm since the weights of MID was set higher than the HFD to avoid the side effects of missing itemsets. When the minimum support threshold is set at 2%, the numbers of deleted items are the same of two algorithms.

7 Conclusions and Future Works

In this paper, the HMU algorithm is proposed for hiding sensitive itemsets in the data sanitization process by reducing the side effects through item deletion. The formulas of two dimensions HFD and MID are defined to evaluate the correlation between the processed transactions and side effects. The weights of two evaluation dimensions of HFD and MID can be set by users' interests. The Experimental results show that the proposed HMU algorithm outperforms the disaggregate algorithm in terms of execution time and number of side effects.

References

1. Agrawal, R., Imielinski, T., Swami, A.: Database mining: A performance perspective. *IEEE Transactions on Knowledge and Data Engineering* 5, 914–925 (1993)
2. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: *The International Conference on Very Large Data Bases*, pp. 487–499 (1994)
3. Atallah, M., Bertino, E., Elmagarmid, A., Ibrahim, M., Verykios, V.: Disclosure limitation of sensitive rules. In: *The Workshop on Knowledge and Data Engineering Exchange*, pp. 45–52 (1999)
4. Agrawal, R., Srikant, R.: Privacy-Preserving Data Mining. In: *ACM SIGMOD International Conference on Management of Data*, pp. 439–450 (2000)
5. Amiri, A.: Dare to share: Protecting sensitive knowledge with data sanitization. *Decision Support Systems* 43, 181–191 (2007)
6. Berkhin, P.: A survey of clustering data mining techniques. *Grouping Multidimensional Data*, 25–71 (2006)
7. Dasseni, E., Verykios, V.S., Elmagarmid, A.K., Bertino, E.: Hiding association rules by using confidence and support. In: *International Workshop on Information Hiding*, pp. 369–383 (2001)
8. Gkoulalas-Divanis, A., Verykios, V.S.: An integer programming approach for frequent itemset hiding. In: *ACM International Conference on Information and Knowledge Management*, pp. 748–757 (2006)
9. Duraiswamy, K., Manjula, D., Maheswari, N.: Advanced approach in sensitive rule hiding. *CCSE Modern Applied Science* 3, 98–107 (2009)
10. Gkoulalas-Divanis, A., Verykios, V.S.: Exact knowledge hiding through database extension. *IEEE Transactions on Knowledge and Data Engineering* 21, 699–713 (2009)
11. Han, J., Fu, Y.: Mining multiple-level association rules in large databases. *IEEE Transactions on Knowledge and Data Engineering* 11, 798–805 (1999)
12. Han, J., Pei, J., Yin, Y., Mao, R.: Mining frequent patterns without candidate generation: a frequent-pattern tree approach. *Data Mining and Knowledge Discovery* 8, 53–87 (2004)
13. Hong, T.P., Lin, C.W., Wu, Y.L.: Incrementally fast updated frequent pattern trees. *Expert Systems with Applications* 34, 2424–2435 (2008)
14. Hong, T.P., Lin, C.W., Yang, K.T., Wang, S.L.: A lattice-based data sanitization approach. *IEEE International Conference on Systems, Man, and Cybernetics*, 2325–2329 (2011)
15. Hong, T.P., Lin, C.W., Yang, K.T., Wang, S.L.: Using TF-IDF to hide sensitive itemsets. *Applied Intelligence* 38, 502–510 (2013)

16. Kotsiantis, S.B.: Supervised machine learning: A review of classification techniques. In: The Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies, pp. 3–24 (2007)
17. Leary, D.E.O.: Knowledge discovery as a threat to database security. *Knowledge Discovery in Databases*, pp. 507–516 (1991)
18. Lin, C.W., Hong, T.P., Lu, W.H.: An effective tree structure for mining high utility itemsets. *Expert Systems with Applications* 38, 7419–7424 (2011)
19. Lin, C.W., Hong, T.P., Chang, C.C., Wang, S.L.: A greedy-based approach for hiding sensitive itemsets by transaction insertion. *Journal of Information Hiding and Multimedia Signal Processing* 4, 201–227 (2013)
20. Lin, C.W., Hong, T.P.: A survey of fuzzy web mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 3, 190–199 (2013)
21. Modi, C.N., Rao, U.P., Patel, D.R.: Maintaining privacy and data quality in privacy preserving association rule mining. In: International Conference on Computing Communication and Networking Technologies, pp. 1–6 (2010)
22. Wu, Y.H., Chiang, C.M., Chen, A.L.P.: Hiding sensitive association rules with limited side effects. *IEEE Transactions on Knowledge and Data Engineering* 19, 29–42 (2007)
23. Zheng, Z., Kohavi, R., Mason, L.: Real world performance of association rule algorithms. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 401–406 (2001)

The Bridge Edge Label Propagation for Overlapping Community Detection in Social Networks

Jui-Le Chen^{1,2}, Jen-Wei Hu¹, and Chu-Sing Yang¹

¹ Institute of Computer and Communication Engineering,
National Cheng Kung University, Taiwan
{q38991051,q38001050,csyang}@mail.ncku.edu.tw

² Department of Computer Science, Tajen University, Taiwan

Abstract. Overlapping community detection aims to discover a set of groups in which each node belongs to at least one group. There are more proposed methods that interest in overlapping community detection to find out the groups which are not necessarily disjoint. In this paper, we propose a modify method that provides the detection results would be the same for each run. The accuracy for experimental result of overlapping community detection is better but not much time consume.

Keywords: Overlapping community, Label Propagation Algorithm.

1 Introduction

Community is formed by members which within a group interact with each other more frequently than with the others outside the group. Community detection aims to discover groups in a network where is given a set that contains members with connection property. However, most of the work has been done on disjoint community detection. Instead of one member just belongs to single community, it is possible for each member to have many communities simultaneously. Overlapping community detection aims to discover a set of groups in which each node belongs to at least one group. For the reason, there are more proposed methods that interest in overlapping community detection to find out the groups which are not necessarily disjoint.

In recent years, there are many kinds of methods which try to identify the overlapping community. (1) Clique percolation method[1] is based on the assumption that a community consists of fully connected sub-graphs and detects overlapping communities by searching for adjacent cliques. It is a popular approach for analyzing the overlapping community structure of networks. (2) Local expansion method[2] used the iterative scan algorithm(IS) to improve . (3) Fuzzy clustering method[3] provided the fuzzy c-mean algorithm. to embed the graph into low dimensionality Euclidean space. (4) Link partitioning method[4] using links instead of nodes to discover communities. (5) Dynamical Algorithms [5] and Speaker-listener Label Propagation Algorithm)[6] which are the label propagation algorithms use labels to discover communities.

The SLPA(Speaker-listener Label Propagation Algorithm)[6] method proposed a method for detecting both individual overlapping nodes and overlapping communities using the underlying network structure alone. SLPA accounts for overlap by allowing each node to possess multiple labels. In each communication step, one node is a speaker (information provider), and the other is a listener (information consumer). each node has a memory of the labels received in the past and takes its content into account to make the current decisions. However, the method although gives the complexity: $O(Tnk)$ with maximum iteration(T), the average node degree(k) and the total number of nodes(n) that towards linear time but not provides the stable detection results.

For the reason, we propose a modify method that provides the detection results would be the same for each run. The accuracy for experimental result of overlapping community detection is better than SLPA but not much time consume.

2 Problem Definition

In this section, we present basic definitions that will be used throughout the paper. Given a network or undirected graph $G = E, V$, V is a set of n nodes and E is a set of m edges. In the case of overlapping community detection, the set of clusters found is called a cover or partition $C = \{c_1, c_2, \dots, c_k\}$, in which a node may belong to more than one cluster.

3 The Proposed Method

In the proposed method, each node is assigned a unique community id as label and maintains a group list with size n , the number of nodes in network. At the first, the group list of each node is initialized with a null label. Then, the following steps are repeated until the maximum iteration T is reached: (a) Each node detects the label of each neighbor, if both are not the same then raising the bridge edge problem and determinates the two nodes are the same community or not. (b) Each node has a group list of the labels to record these labels of its neighbors at each position. Finally, based on the labels in the group list is applied to output the communities by using post-processing and community detection.

The Bridge Edge Problem is defined as the edge connecting two communities. When the bridge edge problem is arising , it means that both node of edge's side would be made decision for belongs to which community. Therefore, the overlap determination can be deal with the kind of the bridge edges. The good judgement is introduce by [7] with average degree for the concept of partition density. The definition of average degree is as follows: where c is a community, $E(c)$ is the number of edges in the community, and $|c|$ is the number of nodes of the community. If adding to the community makes $|AD(c)|$ increase, we suppose that the node contributes to the community.

$$AD(c) = \frac{2 * E(c)}{|c|} \quad (1)$$

Algorithm 1. Bridge Edge Label Propagation Algorithm

```

1: Each node is assigned a unique community id as label.
2: Each node maintains a group list with size  $n$ , the number of nodes in network
   and initialized with a null label.
3: while the termination criterion is not met( $t < T$ ) do
4:   for each Node  $N_i, i = 1$  to  $n$  do
5:     for each neighbor  $N_k$  of  $N_i$  do
6:       if  $(N_i, N_k)$  is not same group then
7:         Calculate  $AD(c_k) = 2 * E(c_k) / |c_k|$ 
8:         Calculate  $AD(c_{i,k}) = 2 * E(c_{i,k}) / |c_{i,k}|$ 
9:         if  $AD(c_k) < AD(c_{i,k})$  then
10:            add Node(K) into C(i)
11:        end if
12:    end if
13:  end for
14: end for
15: end while
16: Post-processing and community detection
17: (a)Detect the transitive property:
18: Transform all of node of  $C(i)$  into  $C(j)$  if  $C(i) \subset C(j)$ 
19: (b)Construct a maximum community:
20: Two communities are adjacent if they share  $k-1$  nodes,  $|C(i)| = k < |C(j)|$ 
21: (c)Make a histogram( $i$ ) but except node( $i$ ) itself for each node's grouplist[ $\hat{i}$ ]:
22: In histogram( $i$ ), find out the frequent  $\geq 2$  then output the community id.

```

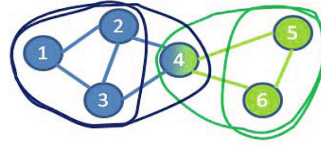
For examples, in Figure 1, there are two Communities 1: {1, 2, 3} and 2: {4, 5, 6} with bridge edges $e(2, 4), e(3, 4)$. At the result, we can find that node 4 is a member of both community 1 and 2, but node 3 and 2 just belong to community 1.

Another example showed in Figure 2: Node 3 is contained in community 1 and Node 4 is in community 2. Because node 3 decreases the $|AD(c)|$ of community 2, while adding into community 2, then node 3 should be just belong to community 1, which is the most reasonable partition. For the same reason, node 5 decreases the $|AD(c)|$ of community 1, while joins to community 1, and then node 5 would be in community 2.

3.1 Complexity Evaluation

The initialization of labels requires $O(n)$, where n is the total number of nodes. Each node has a group list of size n . The operation executed by each node in each iteration. Detects the label of each neighbor is the same or not which requires $O(N * k)$, where k is the average degree of node. If not, the bridge edge problem raised then processing the determination. Each bridge node would find out all edges in the same group which requires $O(2 * k * n_b)$, where n_b is the total number of bridge edges. The n_b would equal to the number of community h so that $O(2 * k * n_b)$ would be the $O(2 * k * h)$. At the result, the complexity

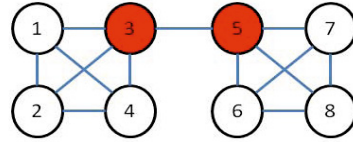
Two communities with bridge edges $e(2,4)$, $e(3,4)$
 (a) Community 1: $\{1,2,3\}$, $AD(c) = 2*(3/3)=2$
 Adding node 4 to community 1:
 $AD(c) = 2*(5/4)=5/2 > 2$, then
 node 4 is also a member in community 1
 (b) Community 2: $\{4,5,6\}$, $AD(c) = 2*(3/3)=2$
 Adding node 2 to community 2:
 $AD(c) = 2*(4/4) = 2$, then
 node 2 does not belongs to community 2.
 Same as node 3 not belongs to community 2



Group List	1	2	3	4	5	6
Node 1	1	1	1	1	0	0
Node 2	1	2	1	1	0	0
Node 3	1	1	3	1	0	0
Node 4	1	1	1	4	5	5
Node 5	0	0	0	5	5	5
Node 6	0	0	0	5	5	6
#freq>2	1	1	1	1,5	5	5

Fig. 1. The overlap determination for network size=6

Two communities with bridge edges $e(3,5)$
 (a) Community 1: $\{1,2,3,4\}$, $AD(c) = 2*(6/4)=3$
 Adding node 5 to community 1:
 $AD(c) = 2*(7/5)=14/5 < 3$, then
 node 5 does not belong to community 1
 (b) Community 2: $\{5,6,7,8\}$, $AD(c) = 2*(6/4)=3$
 Adding node 3 to community 2:
 $AD(c) = 2*(7/6) = 14/5 < 3$, then
 node 3 does not belong to community 2.



Group List	1	2	3	4	5	6	7	8
Node 1	1	1	1	1	0	0	0	0
Node 2	1	2	1	1	0	0	0	0
Node 3	1	1	3	1	0	0	0	0
Node 4	1	1	1	4	5	0	0	0
Node 5	0	0	0	5	5	6	6	6
Node 6	0	0	0	0	6	6	6	6
Node 7	0	0	0	0	6	6	7	6
Node 8	0	0	0	0	6	6	6	8
#freq>2	1	1	1	1	6	6	6	6

Fig. 2. The overlap determination for network size=8

in each iteration would be $O(N * k^2 * h)$ at the worst case(upper bound). The complexity of the other methods for overlapping detection show in the table 1.

Table 1. Complexity for overlapping community detection methods

Algorithm	Time Complexity
Clique Percolation Method(CPM)	$O(n^3)$
Local Expansion	$O(mh)$
Fuzzy Clustering	$O(hn^2)$
Link Partitioning	$O(nk_{max}^2)$
Dynamical Algorithms	$O(Tnk)$
n is the number of nodes, m is the number of edges.	
h is the number of cliques, k_{max} is the highest degree of the n nodes	

4 Experimental Result and Discussion

In this paper, the performance of the proposed algorithm is evaluated by using it to solve the prototype generation in nearest neighbor classification problem. All the experimental results are obtained by running on an IBM X3650 machine with 2.4 GHz Xeon CPU and 16GB of memory using CentOS 6.0 with Linux 2.6.32. Moreover, all the programs are written in C++ and compiled using GNU C++ compiler.

4.1 Parameter Settings and Datasets

To study the behavior of proposed method, we conducted the experiments in synthetic networks. For synthetic random networks, we adopted the widely used LFR[8] benchmark data set. Program source code for generating benchmark data set can be get from <http://sites.google.com/site/andrealancichinetti/files>. The parameters show as table 2.

Table 2. LRF’s synthetic benchmark data set

1. The networks with size $n = 1000$.
2. The average degree is kept at $k = 10$.
3. The node degrees and community sizes are governed by the power laws with exponents 2 and 1;
4. The maximum degree is 50;
5. The community size varies from 20 to 100;
6. The expected fraction of links of a node connecting it to other communities, called the mixing parameter μ , is set to 0.3.
7. O_n defines the number of overlapping nodes is set to 10
8. O_m defines the number of communities to which each overlapping node belongs and varies from 2 to 8 indicating the diversity of overlap.
9. The usage for generating benchmark ./benchmark - N1000 - k10 - t12 - t21 - maxk50 - minc20 - maxc100 - mu0.3 - on500 - om8

4.2 Experimental Results

We focus on proposed method and SLPA method to compare the performance for finding the overlap communities. In total, 5 testing sets as the benchmark were collected and tested. They are listed in Table 3.

Table 3. The performance for SLPA and proposed method

	Execution time(secs)1000/2000/3000/4000/5000	Accuracy rate*
SLPA	25/53/79/113/140	0.3/0.53/0.41/0.37/0.46
Propose Algorithm	27/60/97/132/163	0.64/0.7/0.73/0.69/0.63
*Accuracy rate : all of nodes is in the same community or not / all of nodes		

5 Conclusion

We proposed a improved method for SLPA method to archive an efficient and effective for overlapping community detection. It is important to analyze the information in the social network which can provide more helps to model the overlapping community accurately. To know that how to find out the individual overlapping community with a programmable process.

References

1. Palla, G., Derényi, I., Farkas, I., Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435(7043), 814–818 (2005)
2. Lee, C., Reid, F., McDaid, A., Hurley, N.: Detecting highly overlapping community structure by greedy clique expansion. *arXiv preprint arXiv:1002.1827* (2010)
3. Zhang, S., Wang, R.-S., Zhang, X.-S.: Identification of overlapping community structure in complex networks using fuzzy c-means clustering. *Physica A: Statistical Mechanics and its Applications* 374(1), 483–490 (2007)
4. Ahn, Y.-Y., Bagrow, J.P., Lehmann, S.: Link communities reveal multiscale complexity in networks. *Nature* 466(7307), 761–764 (2010)
5. Gregory, S.: Finding overlapping communities in networks by label propagation. *New Journal of Physics* 12(10), 103018 (2010)
6. Xie, J., Szymanski, B.K.: Towards linear time overlapping community detection in social networks. In: Tan, P.-N., Chawla, S., Ho, C.K., Bailey, J. (eds.) *PAKDD 2012, Part II*. LNCS, vol. 7302, pp. 25–36. Springer, Heidelberg (2012)
7. Cai, Y., Shi, C., Dong, Y., Ke, Q., Wu, B.: A novel genetic algorithm for overlapping community detection. In: Tang, J., King, I., Chen, L., Wang, J. (eds.) *ADMA 2011, Part I*. LNCS, vol. 7120, pp. 97–108. Springer, Heidelberg (2011)
8. Lancichinetti, A., Fortunato, S., Radicchi, F.: Benchmark graphs for testing community detection algorithms. *Physical Review E* 78(4), 046110 (2008)

A New Estimation of Distribution Algorithm to Solve the Multiple Traveling Salesmen Problem with the Minimization of Total Distance

S.H. Chen¹ and Y.H. Chen²

¹ Department of Information Management, Cheng Shiu University. No. 840, Chengcing Rd., Niasong Dist., Kaohsiung City 83347, Taiwan (R.O.C.)

shchen@csu.edu.tw

² Department of Information Science and Applications, Asia University. No. 500, Lioufeng Rd., Wufeng, Taichung 41354, Taiwan (R.O.C.)

chenyh@asia.edu.tw

Abstract. Even though the Estimation of Distribution Algorithms (EDAs) have recently been applied to solve many hard problems, only a few EDAs discussed the in-group optimization problems, such as the multiple traveling salesmen problem (mTSP) studied in this research. These problems include the assignment and sequencing procedures in the same time and to be shown in different forms. As a result, this research proposed an algorithm deal by using the Self-Guided GA together with the Minimum Loading Assignment rule (MLA) to tackle the mTSP. We compare the proposed algorithm against the best direct encoding technique, two-part encoding genetic algorithm, in the experiment on the 33 instances drawn from the well-known TSPLIB. The experimental results show the proposed algorithm is better than the compared algorithm in terms of minimization of the total traveling distance. An interesting result also presents the proposed algorithm would not cause longer traveling distance when we increase the number of salesmen from 3 to 10 persons under the objective of minimization of total traveling distance. This research may suggest the EDAs researcher could employ the MLA rule instead of the direct encoding algorithms.

1 Introduction

Estimation of Distribution Algorithms (EDAs) has been discussed extensively in recent years [4,15,11,13,14]. In particularly, a number of the latest papers on EDAs in solving some NP-hard scheduling problems [12,7,3,9,15,13,10] have shown that EDAs are able to perform effectively. Ceberio et al. [3], in particular, extensively tested 13 famous permutation-based approaches in EDAs on four well-known combinatorial optimization problems. Their paper has provided a good basis for comparison.

Even though EDAs was effective in solving various hard problems, there is a problem that EDA is not discussed extensively. To the best of our knowledge,

only one EDAs proposed by Shim et al. [14] is able to solve in-group optimization problems, such as the Multiple Traveling Salesmen Problems (mTSP) and the Parallel Machine Scheduling Problems (PMSPs) belonged to this category [1]. In-group optimization problems involve the assignment and routing/sequencing procedures in the same time. Take mTSP for example, a number of n cities are assigned to m salesmen and these n cities are visited once by a salesman where $n > m$. It is apparently that this problem is a NP-Hard problem.

Due to there was only a few EDAs could solve the in-group optimization problems, this research proposed an algorithm deal by using the Self-Guided GA [5] together with the Minimum Loading Assignment rule (MLA) to tackle the mTSP. This strategy is called the transformed-based encoding approach instead of the direct encoding. The solution space of the MLA would be only $n!$. We compare the proposed algorithm against the best direct encoding technique, two-part encoding genetic algorithm (TPGA)[2], in the experimental section. It is notable that solution space of the two-part encoding approach is $n! \binom{n-1}{m-1}$. The proposed method MLA, consequently, is better than the two-part encoding technique. A better solution quality is expected when SGGA works with MLA method.

This rest of the paper is organized as follows: We illustrate the core method of the assignment rule in Section 2 which is applied to Self-Guided GA in Section 3. The experimental results are provided in Section 4 and we draw the conclusions in Section 5.

2 Assignment Rule in the mTSP Problems

Given a set of city sequence $\pi_1, \pi_2, \dots, \pi_n$ in π and these cities are not assigned to any salesman yet. This sequence π could be decoded to by assigning the cities to salesmen. That is, the this assignment rule is executed in the fitness function of each chromosome. The rule we called is the minimum loading assignment (MLA) rule. The following pseudo code illustrates the MLA rule.

In the beginning, the first m cities are assigned to the m salesmen and we calculate the objective values of each salesman. The objective function of mTSP would be the total traveling distance or the maximum traveling distance among the salesman. After that, we do the MLA rule iteratively for the unassigned cities. MLA rule assigns the first unassigned city in the sequence π to a salesman when it causes the minimum objective value. This assigned city is removed from the π . This rule is not stopped until there is no city in the π . By using the rule, it means the assigned city could be assigned to a salesman who has the less loading. It also implies that this assigned city might be closed to the last city visited by the salesman so that a far away city would not be considered. Through the MLA rule, it is able to be extended to the parallel machine scheduling problem with setup consideration or the distributed flowshop scheduling problem.

Algorithm 1. Minimum loading assignment rule

Require:

- i : The position of a city in the sequence π
 - $k[i]$: The current number of assigned cities of a salesman i
 - $\Omega_{k[i]}^i$: The
 - 1: $i \leftarrow 1$
 - 2: **while** $i \leq m$ **do**
 - 3: $k[i] \leftarrow 1$
 - 4: $\Omega_{k[i]}^i \leftarrow \pi_i$
 - 5: $i \leftarrow i + 1$
 - 6: $k[i] \leftarrow k[i] + 1$
 - 7: **end while**
 - 8: **while** $i \leq m$ **do**
 - 9: Select a salesman j who could process the π_i with the minimum objective value
 - 10: $\Omega_{k[j]}^j \leftarrow \pi_i$
 - 11: $i \leftarrow i + 1$
 - 12: $k[i] \leftarrow k[i] + 1$
 - 13: **end while**
-

3 Transformed-Based Encoding in Self-Guided Genetic Algorithm

After we introduced the assignment rule in mTSP, this section describes the detail procedures of the Self-guided GA. The benefits of the proposed method are preserving the salient genes of the chromosomes, and exploring and exploiting good searching directions for genetic operators. In addition, since the probabilistic difference provides good neighborhood information, it can serve as a fitness function surrogate. The detailed procedure of the Self-guided GA is described as follows:

Step 1 is the initialization of a population. The sequence of each chromosome is generated randomly.

Step 2 initializes the probability matrix $P(t)$ and the matrix size is $n - by - n$, where n is the problem size. Step 7 builds the probabilistic model $P(t)$ after the selection procedure. In Step 8 and Step 9, $P(t)$ is employed in the self-guided crossover operator and the self-guided mutation operator. The probabilistic model will guide the evolution direction, which is shown in Section 3.2 and Section 3.1. In this research, the two-point central crossover and swap mutation are applied in the crossover and mutation procedures for solving the mTSP under this study.

We explain the proposed algorithm in detail in the following sections. We explain how the probabilistic model guides the crossover and mutation operators.

3.1 Crossover Operator with Probabilistic Model

The idea of Self-Guided Crossover is the same with Self-Guided Mutation, which employs the probability differences of the mating chromosomes by using the

Algorithm 2. MainProcedure of Self-guided GA()

Population: A set of solutions*Generations*: The maximum number of generations*P(t)*: Probabilistic model*t*: Generation index

```

1: Initialize Population
2:  $t \leftarrow 0$ 
3: Initialize  $P(t)$ 
4: while  $t < \text{generations}$  do
5:   EvaluateFitness (Population)
6:   Selection/Elitism(Population)
7:    $P(t+1) \leftarrow \text{BuildingProbabilityModel}(\text{Selected Chromosomes})$ 
8:   Self-Guided Crossover()
9:   Self-Guided Mutation()
10:   $t \leftarrow t + 1$ 
11: end while

```

Eq. 1. By doing so, we could evaluate which chromosome is mated with a parent solution. For the detail description, please refer in [6].

$$\Delta = \Delta_1 - \Delta_2 = \prod_{p \in (CP1 \text{ to } CP2), g=[p]}^n P(\text{Candidate1}_{gp}) - \prod_{p \in (CP1 \text{ to } CP2), g=[p]}^n P(\text{Candidate2}_{gp}). \quad (1)$$

3.2 Mutation Operator with Probabilistic Model

Suppose two jobs i and j are randomly selected and they are located in position a and position b , respectively. p_{ia} and p_{jb} denote job i in position a and job j in position b . After these two jobs are swapped, the new probabilities of the two jobs become p_{ib} and p_{ja} . The probability difference Δ_{ij} is calculated as Eq. 2, which is a partial evaluation of the probability difference because the probability sum of the other jobs remains the same.

$$\begin{aligned} \Delta_{ij} &= P(X') - P(X) \\ &\approx \prod_{p \notin (aorb), g=[p]}^n P_{t+1}(X_{gp}) [(p_{ib}p_{ja}) - (p_{ia}p_{jb})]. \end{aligned} \quad (2)$$

Now that the part of $\prod_{p \notin (aorb), g=[p]}^n P_{t+1}(X_{gp})$ is always ≥ 0 , it can be subtracted and Eq. 2 is simplified as follows:

$$\Delta_{ij} = (p_{ib}p_{ja}) - (p_{ia}p_{jb}). \quad (3)$$

$$\Delta_{ij} = (p_{ib} + p_{ja}) - (p_{ia} + p_{jb}). \quad (4)$$

If Δ_{ij} is positive, it implies that one gene or both genes might move to a promising area. On the other hand, when Δ_{ij} is negative, the implication is that at least one gene moves to an inferior position.

On the basis of the probabilistic differences, it is natural to consider different choices of swapping points during the mutation procedure. A parameter TM is introduced for the self-guided mutation operator, which denotes the number of tournaments in comparing the probability differences among the TM choices in swap mutation. Basically, $TM \geq 2$ while $TM = 1$ implies that the mutation operator mutates the genes directly without comparing the probability differences among the different TM choices.

When $TM = 2$, suppose the other alternative is that two jobs m and n are located in position c and position d , respectively. The probability difference of exchanging jobs m and n is:

$$\Delta_{mn} = (p_{md} + p_{nc}) - (p_{mc} + p_{nd}). \quad (5)$$

After Δ_{ij} and Δ_{mn} are obtained, the difference between the two alternatives is as follows:

$$\Delta = \Delta_{ij} - \Delta_{mn}. \quad (6)$$

If $\Delta < 0$, the contribution of swapping job m and n is better, so we swap job m and n . Otherwise, jobs i and j are swapped. Consequently, the option of a larger probability difference is selected and the corresponding two jobs are swapped. By observing the probability difference Δ , the self-guided mutation operator exploits the solution space to enhance the solution quality and prevent destroying some dominant genes in a chromosome. Moreover, the main procedure of the self-guided mutation is Eq. 6, where the time-complexity is only a constant after the probabilistic model is employed. This approach proves to work efficiently.

To conclude, the Self-guided GA is obviously different from the previous EDAs. Firstly, the algorithm utilizes the transformed-based encoding instead of using the direct encoding used by Shim et al. [14]. Secondly, the proposed algorithm explicitly samples new solutions without using the crossover and mutation operators. The Self-guided GA embeds the probabilistic model in the crossover and mutation operators to explore and exploit the solution space. Most important of all, the algorithm works more efficiently than previous EDAs [14] in solving the mTSP because the time-complexity is $O(n)$ whereas the previous EDAs needs $O(n^2)$ time.

4 Experimental Results of the Proposed Algorithm

4.1 Experiment Settings

We conducted extensive computational experiments to evaluate the performance of Self-guided GA together with the MLA rule in solving the mTSP. The proposed algorithm was compared with the benchmark encoding algorithm, Two-Part chromosome GA, from the literature [2]. In addition, we employ the genetic

operators and parameter settings of Two-Part chromosome genetic algorithm suggested Chen and Chen [8]. The genetic operators are the two-point crossover operator and the swap mutation operator. As a result, it ensures we do a fair comparison between the proposed algorithm with the benchmark encoding algorithm. Besides, a standard genetic algorithm (SGA) also applies the MLA rule which could show the performance enhanced by the assignment rule proposed by this research.

The objective function is to minimize the total traveling distance which is shown in Section 4.2. We implemented the algorithms in Java 2 on a Amazon EC2 with the Windows 2008 server (8-cores CPU). Across all the experiments, we replicated each instance 30 times on the 33 instances from the well-known TSPLIB. We assume the first city of each instance is the home-depot. The size of these instances is from 48 to 400. The number of salesmen is ranging from 2, 3, 5, 10, and 20. As a result, we conduct extensive experiments to evaluate the proposed algorithm under different circumstances.

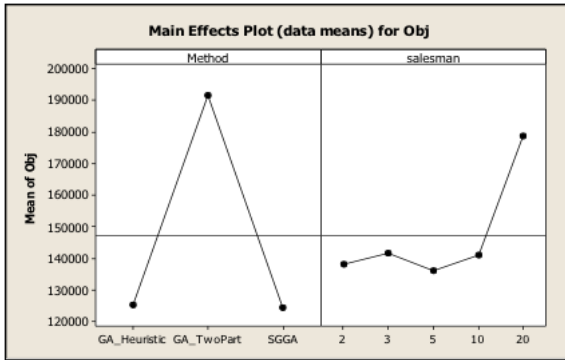


Fig. 1. Main effects plot on the total traveling distance of the compared algorithms

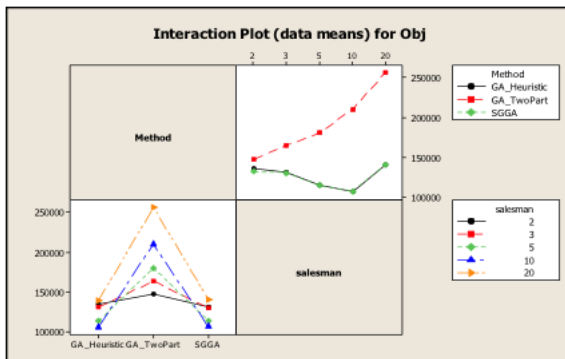


Fig. 2. Intreaction plot on the total traveling distance of the compared algorithms

Table 1. Total Distance

Instance	<i>GA heuristic</i>			<i>GA twoPart</i>			SGGA		
	Minimum	Mean	Maximum	Minimum	Mean	Maximum	Minimum	Mean	Maximum
att48	39873	67412	119821	41382	80572	142234	38279	67408	121381
berlin52	9194	13307	20647	9911	15078	23586	8773	13397	21140
bier127	220310	257767	329280	238090	318724	433197	214928	257983	311903
ch130	12781	16173	19262	15896	22951	32313	12020	15959	19035
eil101	1024.4	1251.1	1625.4	1170.4	1734.2	2441.7	910.5	1230.2	1624.1
eil51	497.2	726.6	1195.1	566	896.7	1450.8	483.7	726.2	1186
eil76	771.2	1036.2	1640.5	840.8	1365.6	2151.2	697.2	1030.3	1610.3
gr96	911.3	1325.6	2260.3	1103.9	1834.4	2968.2	893	1325.3	2257.7
kroa150	62801	80353	104886	84155	118336	164472	60914	77752	93631
kroa200	90234	114171	145046	122311	164758	223643	86536	112027	141175
kroB100	39466	52960	74861	49619	78483	118140	39213	52689	78913
kroB150	62532	80549	93700	81609	122203	171691	63013	79185	97548
kroB200	85875	114978	149151	117037	162784	220721	88753	113813	142722
kroC100	37895	53097	77495	55150	81016	128239	37420	51764	78388
kroD100	40864	54795	82342	51506	80103	129610	38858	53964	81408
kroE100	39623	55741	80784	52077	82172	128825	38145	55008	85800
lin105	30255	45563	79901	37418	65266	107157	26922	45533	77804
lin318	182437	243339	322419	251134	328117	447096	183700	243829	318757
mTSP100	40596	54269	81858	50939	80380	123128	39229	54107	83089
mTSP150	76189	96634	116498	101690	138019	184927	76317	94583	118046
mTSP51	467.2	725.2	1188	558.9	897.7	1436.8	475.1	727.6	1198.2
pr124	133373	201905	298646	199047	328870	507277	134543	198651	296597
pr136	223163	282285	383137	273500	419743	634173	200711	281343	390805
pr144	194333	253771	327206	265350	399668	587282	198892	251187	323671
pr152	207562	309617	409260	332865	524397	790537	217451	304029	403995
pr226	329781	519456	753744	588036	842705	1223229	343841	517697	728401
pr264	172047	244558	386775	360147	504250	698832	173869	239184	314209
pr299	218292	283892	361192	314827	413047	585709	220328	282296	357225
pr76	167847	251580	434977	185337	334205	560341	148840	252445	440088
rat195	7200.4	8775.9	10212.1	8563	12055	16873	7059.3	8667.7	10285.1
rat99	2144.1	3418.9	5939.7	2768	4606	7707	2013.9	3392.5	5924
rd400	53411	87481	126827	91499	111470	135460	53204	88769	122815
st70	939.3	1497.3	2594	1142.5	2005.8	3315.9	929.4	1495.7	2641.4
tsp225	13615	16933	20478	17201	23387	32360	13417	16975	20124

4.2 Results of the Total Traveling Distance

This objective evaluates the total distances travelled by the m salesmen. It reflects the total cost of the assignment. Fig. 1 shows the main effects plot on the method comparison and the differences of the number of salesmen we assign. This figure clearly illustrates the SGGA and SGA (denoted $GA_{Heuristic}$) are better than the Two-Part encoding GA (named $GA_{TwoPart}$). It means the MLA rule, i.e. the transformed-based method, could be a promising approach which is better than the direct encoding method. Then, when the number of salesmen increased, especially there are 20 salesmen could be assigned, the total distance is increased greatly. As a result, it implies the inefficiency if we request too many salesmen in terms of the managerial perspective.

Fig. 2 depicts the interaction plot between the factor method and the number of salesmen. It might be interesting to see the SGGA and SGA that do not yield the longer total traveling distance when the number of salesmen increased from two to 10 salesmen. However, Two-Part encoding GA may suffer the pain of the number of salesmen increased. This figure could distinguish the effectiveness for the transform-based rule to the direct encoding method. Finally, if a manager would like to determine how many salesmen is required, the lowest total traveling distance would be ten according to this interaction plot.

Finally, the detail result of the three compared algorithms is shown in Table 1.

5 Conclusions

This study solve the in-group optimization problems which is rarely solved by the EDAs. A new EDAs SGGA was proposed, which works with the MLA rule together. In addition, because the MLA rule is classified in the category of transform-based encoding, the proposed algorithm is compared with the two-part encoding GA which is the best direct encoding strategy so far. We evaluate these algorithm by solving the mTSP problem under 33 instances drawn from TSPLIB. The experimental results show the SGGA with the MLA rule outperforms the Two-Part encoding GA in both the total traveling distance and the maximum traveling objectives. It reveals the proposed algorithm is capable for solving the mTSP problem well. In addition, the MLA rule is also effective and could be applied on some GAs that originally designed for the permutation type problems. As a result, this research provides an insightful results for the researchers who are doing the scheduling problems and could move toward the in-group optimization problems.

References

1. Bektas, T.: The multiple traveling salesman problem: an overview of formulations and solution procedures. *Omega* 34(3), 209–219 (2006)
2. Carter, A.E., Ragsdale, C.T.: A new approach to solving the multiple traveling salesperson problem using genetic algorithms. *European Journal of Operational Research* 175(1), 246–257 (2006)

3. Ceberio, J., Irurozki, A.M.E., Lozano, J.: A review on Estimation of Distribution Algorithms in Permutation-based Combinatorial Optimization Problems. Accepted by Progress in Artificial Intelligence (2011)
4. Chang, P.C., Hsieh, J.C., Chen, S.H., Lin, J.L., Huang, W.H.: Artificial chromosomes embedded in genetic algorithm for a chip resistor scheduling problem in minimizing the makespan. *Expert Systems With Applications* 36(3 Pt. 2), 7135–7141 (2009)
5. Chen, S.H., Chang, P.C., Cheng, T.C.E., Zhang, Q.: A self-guided genetic algorithm for permutation flowshop scheduling problems. *Computers & Operations Research* 39(7), 1450–1457 (2012)
6. Chen, S.H., Chang, P.C., Edwin Cheng, T.C., Zhang, Q.: A self-guided genetic algorithm for permutation flowshop scheduling problems. *Computers & Operations Research* 39(7), 1450–1457 (2012)
7. Chen, S.H., Chang, P.C., Zhang, Q., Wang, C.B.: A guided memetic algorithm with probabilistic models. *International Journal of Innovative Computing, Information and Control* 5(12), 4753–4764 (2009)
8. Chen, S.H., Chen, M.C.: Operators of the two-part encoding genetic algorithm in solving the multiple traveling salesmen problem. In: *The 2011 Conference on Technologies and Applications of Artificial Intelligence, TAAI 2011* (2011)
9. Chen, S.H., Chen, M.C., Chang, P.C., Chen, Y.M.: Ea/g-ga for single machine scheduling problems with earliness/tardiness costs. *Entropy* 13(6), 1152–1169 (2011)
10. Chen, S.-H., Chen, M.-C.: Addressing the advantages of using ensemble probabilistic models in estimation of distribution algorithms for scheduling problems. *International Journal of Production Economics* 141(1), 24–33 (2013)
11. Hauschild, M., Pelikan, M.: An introduction and survey of estimation of distribution algorithms. Accepted by *Swarm and Evolutionary Computation* (2011)
12. Jarboui, B., Eddaly, M., Siarry, P.: An estimation of distribution algorithm for minimizing the total flowtime in permutation flowshop scheduling problems. *Computers & Operations Research* 36(9), 2638–2646 (2009)
13. Pan, Q., Ruiz, R.: An estimation of distribution algorithm for lot-streaming flow shop problems with setup times. *Omega* 40(2), 166–180 (2012)
14. Shim, V.A., Tan, K., Cheong, C.: A hybrid estimation of distribution algorithm with decomposition for solving the multiobjective multiple traveling salesman problem. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 42(5), 682–691 (2012)
15. Zhang, Y., Li, X.: Estimation of distribution algorithm for permutation flow shops with total flowtime minimization. *Computers & Industrial Engineering* 60(4), 706–718 (2011)

Subspace Learning with Enriched Databases Using Symmetry

Konstantinos Papachristou, Anastasios Tefas, and Ioannis Pitas

Aristotle University of Thessaloniki,
Department of Informatics,
Box 451, 54124 Thessaloniki, Greece
{kpapaxristou,tefas,pitas}@aiaa.csd.auth.gr

Abstract. Principal Component Analysis and Linear Discriminant Analysis are of the most known subspace learning techniques. In this paper, a way for training set enrichment is proposed in order to improve the performance of the subspace learning techniques by exploiting the a-priori knowledge that many types of data are symmetric. Experiments on artificial, facial expression recognition, face recognition and object categorization databases denote the robustness of the proposed approach.

Keywords: Subspace Learning, Data Enrichment, Symmetry, Principal Component Analysis, Linear Discriminant Analysis.

1 Introduction

Everyday, a vast amount of images and videos are available from many sources, resulting in the need to handle and use this information intelligently by many systems such as robotics, multimedia retrieval and recognition (face, object, etc). This means that image processing methods are a key field in computer vision applications. Many of these methods exploit subspace learning techniques which have been employed in many computer vision and pattern recognition tasks [1,2]. Such techniques calculate projection vectors in order to reduce the data dimensionality, maintaining the meaningful information and, thus, they can be employed for dimensionality reduction, data visualization and compression, as well as as a main preprocessing step in classification and clustering methods. Some of them are unsupervised, such as Principal Component Analysis (PCA) [3], Independent Component Analysis [4], Locality Preserving Projections [5] and Non-negative Matrix Factorization [6]. Another category of SL techniques is supervised and uses the class label information of data, e.g., Linear Discriminant Analysis (LDA) [7], Discriminant Non-negative Matrix Factorization [8], Clustering based Discriminant Analysis [9] and Subclass Discriminant Analysis [10].

The aforementioned techniques do not work well when the available samples are not truly representative of the corresponding patterns. Our aim is to propose a training set enrichment approach in order to produce more representative

training sets and, therefore, to improve the performance of subspace learning techniques by adding the symmetric version of each sample. This approach is based on the fact that symmetry is a main characteristic of several data types, such as faces, objects, etc.

The remainder of this paper is organized as follow. In Section 2, the subspace learning techniques, namely PCA and LDA, are briefly described. In Section 3, the proposed approach for improving the robustness of the subspace learning techniques using the symmetric versions of images are presented. In Section 4, we present experiments conducted in order to evaluate the proposed approach. Finally, conclusions are drawn in Section 5.

2 Subspace Learning Techniques

In this section, we provide a brief review of well known subspace learning techniques Principal Component Analysis in subsection 2.1, LDA in subsection 2.2 and their combination in subsection 2.3. In the following, we will consider the set $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ to be the sample images $\mathbf{x}_i \in \mathcal{R}^{m \times 1}$ in vectorized form, while the projection vectors are denoted by $\mathbf{w} \in \mathcal{R}^{m \times 1}$. The total number of samples in the dataset, the total number of classes and the mean vector of the entire data set are denoted by N , c and $\boldsymbol{\mu}$, respectively. The initial dimensionality of the samples is denoted by m , while the dimensionality of the projection space is denoted by m' .

2.1 Principal Component Analysis

PCA tries to find projection vectors \mathbf{w} that maximize the variance of the projected samples $y_i = \mathbf{w}^T \mathbf{x}_i$, for better representation. If we define the total scatter matrix \mathbf{S}_T as:

$$\mathbf{S}_T = \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T, \quad (1)$$

the objective of PCA is to find the transformation matrix $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{m'}]$ that maximizes the trace of \mathbf{S}_T :

$$J(\mathbf{W}) = \arg \max_{\mathbf{W}} \text{tr}[\mathbf{W}^T \mathbf{S}_T \mathbf{W}]. \quad (2)$$

The solution of (2) is given by the solution of following generalized eigenvalue decomposition problem:

$$\mathbf{S}_T \cdot \mathbf{w} = \lambda \cdot \mathbf{w} \quad (3)$$

keeping the m' eigenvectors of \mathbf{S}_T that correspond to the m' largest eigenvalues. We can choose m' such that the sum of the m' largest eigenvalues is more than a percentage $P\%$ of the sum of the total eigenvalues.

2.2 Linear Discriminant Analysis

LDA determines projection vectors \mathbf{w} so that the classes of the samples are well discriminated. For this reason, the between-class scatter matrix:

$$\mathbf{S}_B^{LDA} = \sum_{i=1}^c (\boldsymbol{\mu}_i - \boldsymbol{\mu}) (\boldsymbol{\mu}_i - \boldsymbol{\mu})^T \quad (4)$$

and the within-class scatter matrix:

$$\mathbf{S}_W^{LDA} = \sum_{i=1}^c \sum_{k=1}^{n_i} (\mathbf{x}_k^i - \boldsymbol{\mu}_i) (\mathbf{x}_k^i - \boldsymbol{\mu}_i)^T, \quad (5)$$

are defined, where \mathbf{x}_k^i is the k -th sample in the class i and, $\boldsymbol{\mu}_i$, n_i are the mean vector and the number of samples in class i , respectively.

The objective of LDA is to find the transformation matrix \mathbf{W} that maximizes the ratio of the trace of the between-class scatter to the trace of the within-class scatter matrix:

$$J(\mathbf{W}) = \arg \max_{\mathbf{W}} \frac{\text{tr}[\mathbf{W}^T \mathbf{S}_B^{LDA} \mathbf{W}]}{\text{tr}[\mathbf{W}^T \mathbf{S}_W^{LDA} \mathbf{W}]} \quad (6)$$

The solution of (6) is approximated [19] by the following generalized eigenvalue decomposition problem:

$$\mathbf{S}_B^{LDA} \cdot \mathbf{w} = \lambda \cdot \mathbf{S}_W^{LDA} \cdot \mathbf{w}, \quad (7)$$

by keeping the m' eigenvectors that correspond to the m' largest eigenvalues. Because \mathbf{S}_B^{LDA} is the sum of c matrices in (Equation 4) of rank one or less and only $c-1$ of these are independent, the maximum number of nonzero eigenvalues is equal to $c-1$. Consequently, the upper bound on m' is $c-1$.

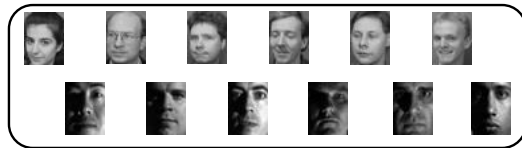
2.3 Principal Component Analysis plus Linear Discriminant Analysis

LDA is very prone to the “small sample size” problem [1]. This problem occurs when the number of samples is smaller than the dimensionality of the samples. As a result, the matrix \mathbf{S}_W^{LDA} may become singular, and solving the generalized eigenvalue decomposition problem (7) may result to irregular discriminant projection vectors.

In order to overcome the above problem, an alternative technique has been proposed [11], which consists of two steps. In the first step, the samples are projected to a subspace of dimensionality lower than $N-l$ using PCA, where l denotes the number of classes for LDA technique, so that \mathbf{S}_W^{LDA} become non-singular. In the second step, the matrices \mathbf{S}_B^{LDA} and \mathbf{S}_W^{LDA} are calculated by using the data representations in the PCA space. Finally, LDA is applied for the determination of regular projection vectors.

3 Proposed Approach

The above mentioned subspace learning techniques are rather sensitive when the training set consists of a small number of samples, resulting in a bad pattern learning and generalization. For example, as illustrated in Figure 1(a), the training set of a face recognition problem may be comprised of frontal and slightly left pose face images or face images taken with a specific light position (right). This fact can lead to a poor pattern representation by applying a subspace learning technique. A possible solution to address this problem is the enrichment of the training set by adding the symmetric version of each sample based on the symmetry property of the face in order to produce a training set which will better represent a symmetric pattern. Indeed, the application of this way of database enrichment to the images of Figure 1(a) leads to forming an enriched training set which represents better the faces of persons, as shown in Figure 1(b). These image have been inverted with respect to the vertical axis. Similarly, we can apply a corresponding training set enrichment by inverting the images with respect to the horizontal axis or to any directional axis.



(a)



(b)

Fig. 1. Training set example consisting of (a) the original samples, and (b) both original samples and their symmetric versions

To highlight the effectiveness of the proposed training set enrichment approach in the subspace learning techniques, we designed two artificial data problems for PCA and LDA, respectively. Figure 2 illustrates the result of PCA for a symmetric artificial data problem, where the real symmetric pattern is defined

by an ellipse, while the available samples are represented by crosses. As can be seen, the available samples do not correspond to a representative subset of the pattern. As a result, the PCA projection line, maximizing the samples variance, is not suitable for the real symmetric pattern. On the contrary, it is obvious that PCA results to a better projection line when the symmetric versions of samples are used.

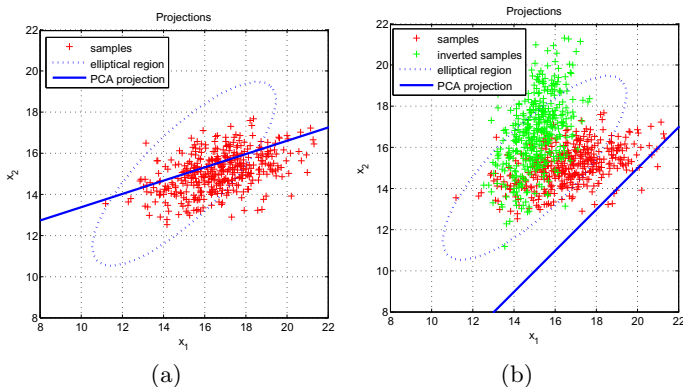


Fig. 2. PCA projection lines using (a) the original samples, and (b) both original samples and their symmetric versions

Correspondingly, we designed an artificial two-class data problem, in which the available samples of the two classes are represented by crosses and circles, respectively. As it can be easily observed in Figure 3, LDA is able to find a projection line, which optimally separates both the available samples and the real symmetric patterns using the enriched training set compared to using the available samples only.

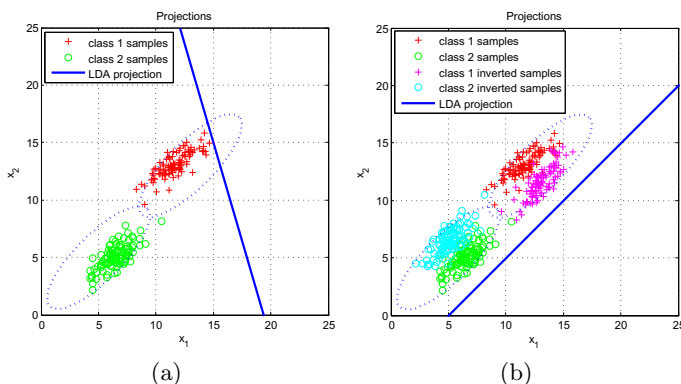


Fig. 3. LDA projection lines using (a) the original samples, and (b) both original samples and their symmetric versions

4 Experiments

In this section, an experimental evaluation of the proposed approach on real-databases for facial expression recognition, face recognition and object categorization is presented. We conducted two series of experiments. In the first one, the training and testing set consist of the original images of databases. In the second one, the original images and their symmetric versions were used to form the training set, while the testing set consists of the original images. In all the experiments, we applied a subspace learning technique, namely PCA, LDA and PCA+LDA, to the training set and the samples are projected into the corresponding subspace. The new dimensionality of PCA has been defined by maintaining the 99% of the total eigenvalue sum of the training set energy, while in LDA technique the new dimensionality was $c-1$, where c is the number of classes. Finally, the projected samples were classified using the Nearest Centroid (NC), and k-Nearest Neighbor (kNN) classifiers. kNN was used for $k = 1, 3, 5, 7, 9, 11$. In all classifiers, the Euclidean distance measure is adopted. The results of our experiments on facial expression recognition, face recognition and object categorization are presented in subsections 4.1, 4.2 and 4.3, respectively.

4.1 Experiments on Facial Expression Recognition

The COHN-KANADE [12] and JAFFE [13] face databases were used in our experiments for facial expression recognition. Each facial image belongs to one of the following seven facial expressions: anger, disgust, happiness, fear, sadness, surprise and neutral. The COHN-KANADE database contains 210 subjects of age between 18 and 50 years. We used 35 images of each facial expression. The JAFFE database contains 213 images depicting 10 Japanese female subjects. 3 images per subject of each facial expression were used in our experiments. All facial images were cropped to include only the subject’s facial region. The cropped face images were resized to 30×40 pixels (where 30 and 40 are the columns and rows of the image, respectively). In Figure 4, a cropped facial image for all facial expressions of the COHN-KANADE and JAFFE databases is shown, respectively.

The application of LDA technique on the above databases encounters computational difficulties due to the “small sample size” problem. To estimate the recognition accuracy, we used the 5-fold cross validation procedure by dividing each database into 5 non-overlapping subsets. Each experiment included five training-test procedures (folds), where in each fold, the techniques were trained by using 4 subsets and testing was performed on the remaining subset. Recognition accuracy was measured by using the mean classification rate over all five folds. For the COHN-KANADE experiments, each subset contained 20% of the facial images for each class based on random selection. For the JAFFE database, we performed person-independent experiments: each subset contained the entire set of the facial images from 20% of the persons. Thus, the facial images of each person were either in the training or in the test set. The results obtained for the COHN-KANADE and JAFFE databases, are shown in Table 1, where the best

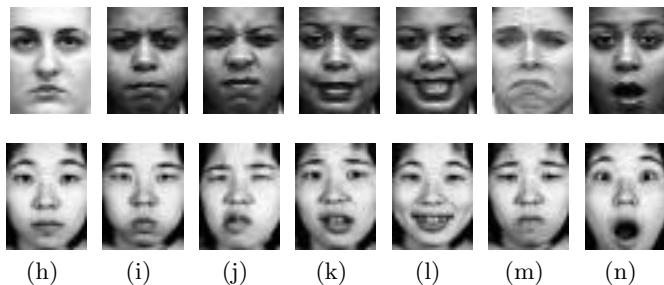


Fig. 4. A cropped image for all facial expressions of the Cohn-Kanade (first row) and JAFFE (second row) databases: (a) neutral, (b) angry, (c) disgusted, (d) feared, (e) happy, (f) sad, and (g) surprised

results are shown in bold. As it can be seen, an improvement in the performance is observed in the majority of the cases after the enrichment with symmetric images. Thus, such an approach can be used in order to improve the performance of subspace learning techniques.

Table 1. COHN-KANADE and JAFFE 5-fold cross validation accuracy rates

technique	COHN-KANADE		JAFFE	
	Original	Enriched	Original	Enriched
PCA	33.88	35.10	38.10	40.48
PCA+LDA	68.98	70.61	51.90	50.00

4.2 Experiments on Face Recognition

We used the ORL [14], AR [18,16] and Extended YALE-B [15,17] face databases in our experiments for face recognition. The ORL database contains 400 images of 40 distinct persons (10 images each). The images were captured at different times and with different variations (lighting, position). The AR database contains over 4000 color images corresponding to 70 men’s and 56 women’s faces. The images were taken in frontal position with different facial expressions, illumination conditions and occlusions. Each person contains 26 images capturing in two recording sessions. The Extended YALE-B database contains images of 38 persons in 9 poses and under 64 illumination conditions. The frontal cropped images were used only, in this work. All images were resized to 30×40 pixels, in our experiments. Some example facial images from the ORL, the AR and the Extended YALE-B databases are displayed in Figure 5.

For the above databases, the 20% of images per person were randomly selected for training and the remaining images were used for testing. The direct application of the LDA technique in all the databases was impossible because of the “small sample size” problem. The results obtained for the ORL, AR and

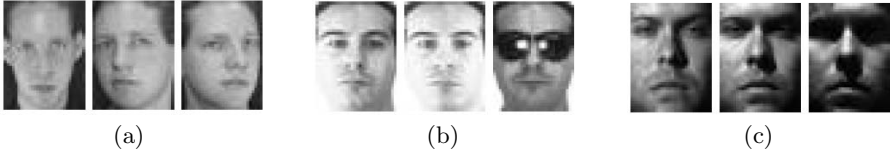


Fig. 5. Sample images from the (a) ORL, (b) AR and (c) Extended YALE-B databases

Extended YALE-B databases, are illustrated in Table 2. As can be seen, in all the cases, a better recognition accuracy is achieved when training set is enriched with the symmetric versions of the original images. Therefore, we can conclude that for symmetric data (such as a human face) the proposed way of enriching databases achieves better data representation and overcomes the poor representation using both the original images and their symmetric versions.

Table 2. ORL, AR and Extended YALE-B Accuracy Rates

technique	ORL		AR		Extended YALE-B	
	Original	Enriched	Original	Enriched	Original	Enriched
PCA	81.88	83.75	27.81	31.95	55.06	55.57
PCA+LDA	80.94	85.63	48.86	53.81	81.53	82.77

4.3 Experiments on Object Categorization

In the experiments on object categorization we used the ETH-80 [20] database. It contains images from eight categories: apple, pear, tomato, cow, horse, dog, cup and car. For each category there are images of ten different objects. Each object has been captured by 41 different views. The images were resized to 32×32 pixels.

Table 3. ETH-80 5-fold cross validation accuracy rates

technique	Original	Enriched
PCA	85.43	85.67
LDA	74.88	81.52
PCA+LDA	85.00	84.58

We evaluated the performance of the proposed techniques using the 5-fold cross validation procedure. Specifically, images of each object were either in the training set or the test set. The results are shown in Table 3. As can be seen, after the enrichment with symmetric images, an improvement in the performance is observed (PCA and LDA cases). On the other hand, when PCA is applied first, the projected samples are not symmetric in PCA space and, therefore, the symmetric versions of the samples do not affect on the performance of LDA.

5 Conclusions

Subspace Learning techniques have been a useful tool in many applications. In this paper, we proposed an enrichment approach of the training set by adding the symmetric versions of the available samples in problems where the patterns are symmetric, for example facial expression and face recognition ones. The experiments on relevant databases and artificial data highlight that a major improvement is achieved when using subspace learning combined with symmetric enrichment training sets.

Acknowledgment. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 316564 (IMPART) and was partially supported by the COST Action IC1106. This publication reflects only the authors views. The European Union is not liable for any use that may be made of the information contained therein.

References

1. Fukunaga, K.: Introduction to Statistical Pattern Recognition, 2nd edn. Academic Press Professional (1990)
2. Jain, A., Duin, R., Mao, J.: Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(1), 4–37 (2000)
3. Jolliffe, I.: Principal Component Analysis, 2nd edn. Springer (2002)
4. Lee, T.-W.: Independent Component Analysis: Theory and Applications. Kluwer Academic Publishers (1998)
5. He, X., Niyogi, P.: Locality preserving projections. In: *Advances in Neural Information Processing Systems*, vol. 16, pp. 153–160 (2003)
6. Lee, D., Seung, H.: Learning the parts of objects by non-negative matrix factorization. *Nature* 401(6755), 788–791 (1999)
7. Fisher, R.A.: The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7(7), 179–188 (1936)
8. Zafeiriou, S., Tefas, A., Buciu, I., Pitas, I.: Exploiting discriminant information in non-negative matrix factorization with application to frontal face verification. *IEEE Transactions on Neural Networks* 17(3), 683–695 (2006)
9. Chen, X.-W., Huang, T.: Facial expression recognition: a clustering-based approach. *Pattern Recognition Letters* 24(9-10), 1295–1302 (2003)
10. Zhu, M., Martínez, A.: Subclass discriminant analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(8), 1274–1286 (2006)
11. Swets, D., Weng, J.: Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(8), 831–836 (1996)
12. Kanade, T., Tian, Y., Cohn, J.: Comprehensive database for facial expression analysis. In: *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 46–53. IEEE Computer Society (2000)
13. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with Gabor wavelets. In: *Proceedings of the 3rd International Conference on Face and Gesture Recognition*, pp. 200–205. IEEE Computer Society (1998)

14. Samaria, F., Harter, A.: Parameterisation of a stochastic model for human face identification. In: Proceedings of 2nd IEEE Workshop on Applications of Computer Vision, pp. 138–142. IEEE Computer Society (1994)
15. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(6), 643–660 (2001)
16. Martínez, A., Kak, A.: PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(2), 228–233 (2001)
17. Lee, K.-C., Ho, J., Kriegman, D.: Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(5), 684–698 (2005)
18. Martínez, A., Benavente, R.: The AR face database. CVC Technical Report, vol. 24 (1998)
19. Wang, H., Yan, S., Xu, D., Tang, X., Huang, T.: Trace ratio vs. ratio trace for dimensionality reduction. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
20. Leibe, B., Schiele, B.: Analyzing Appearance and Contour Based Methods for Object Categorization. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 409–415. IEEE Computer Society (2003)

Image Categorization Using Macro and Micro Sense Visual Vocabulary

Chang-Ming Kuo, Chi-Kao Chang, Nai-Chung Yang,
Chung-Ming Kuo*, and Yu-Ming Chen

Department of Information Engineering, I-Shou University
Dashu, Kaohsiung, Taiwan
kuocm@isu.edu.tw

Abstract. Visual vocabulary representation approach has been successfully applied to many multimedia and vision applications, including visual recognition, image retrieval, and scene modeling/categorization. The idea behind the visual vocabulary representation is that an image can be represented by visual words, a collection of local features of images. In this work, we will develop a new scheme for the construction of visual vocabulary based on the analysis of visual word contents. By considering the content homogeneity of visual words, we design a visual vocabulary which contains macro-sense and micro-sense visual words. The two types of visual words are appropriately further combined to describe an image effectively. We also apply the visual vocabulary to construct image categorization system. The performance evaluation for the system indicates that the proposed visual vocabulary achieves promising results.

Keywords: Visual words, Macro-sense, Micro-sense, Image categorization.

1 Introduction

Recently, visual vocabulary (or bag-of-visual words) representation approach has been successfully applied to many multimedia and vision applications [1]-[6], including visual recognition, image retrieval, scene modeling/categorization [7]-[10] etc., because of the richness of local information and robustness to occlusions, geometric deformations and illumination variations. Generally, the visual vocabulary is generated by training on a set of image blocks or feature points. Each training sample is described by a feature vector [1][3], and then grouped using a clustering algorithm such as K -means. A visual word is then defined as the center of a cluster. So, visual words can be viewed as the representative parts of the training images or objects such as human's eyes or car wheels. The visual word representation for image is similar to word representation for text, in which same words correspond to same object contents. The visual words will contain all important local information of images and can effectively describe the images if the number of visual words is

* Corresponding author.

large enough. Thus visual word representation is very robust and efficient approach for image categorization.

In this paper, considering the characteristics of visual words, we will try to construct a new block-based visual vocabulary for the applications of image categorization. By taking the inhomogeneous and incomplete content of visual words into account, we design a visual vocabulary which contains macro-sense and micro-sense visual words. The two types of visual words are appropriately further combined to describe an image effectively.

The paper is organized as follows. In section 2, the proposed visual vocabulary construction method is discussed. The performance evaluation is presented in Section 3, and finally the conclusion is drawn in Section 4.

2 Construction of Visual Vocabulary for Image Description and Categorization

The overall structure of the proposed image description scheme is illustrated in Fig. 1. It includes three major components: visual vocabulary construction, image description and similarity measure.

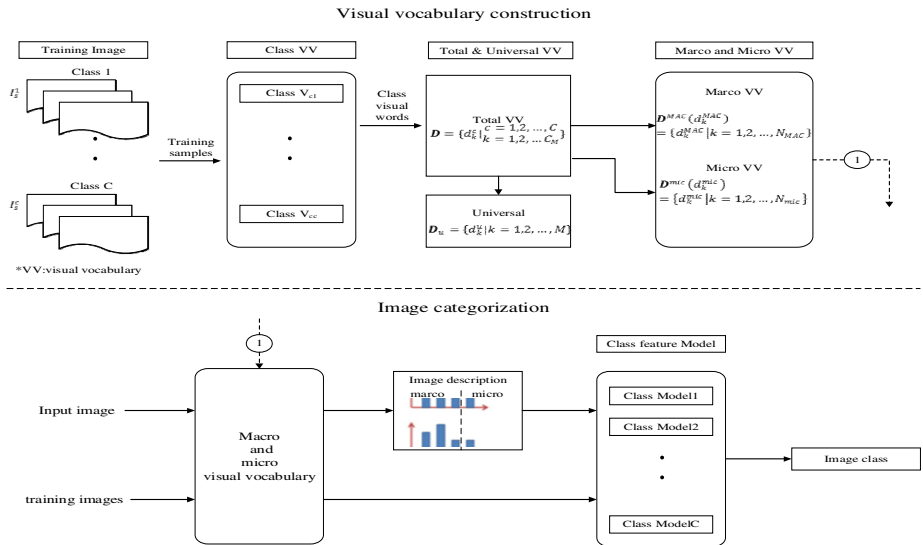


Fig. 1. The overall structure of proposed image description scheme

The Construction of Macro and Micro Sense Visual Vocabulary

To reduce the redundancy to obtain a good visual vocabulary, we design a merge procedure to train a class vocabulary. Instead of using conventional clustering such as K-means which needs to predefine the number of clusters, in our work we use the block similarity measure, which is calculated by the their Euclidean distance, to train

visual words dynamically; hence the class with high details will collect more visual words than that of smooth class. To describe the construction of total vocabulary, the image description based on visual words is introduced below. Assume that an input image is partitioned into N blocks and each block is labeled by the index of nearest visual words in the class vocabulary with size of C_M ; that is

$$L(B_s^c(n)) = k = \underset{k}{\operatorname{argmin}} (\|B_s^c(n) - d_k\|), k = 1, \dots, C_M, n = 1, \dots, N \quad (1)$$

where $L(\bullet)$ is function of labeling, $B_s^c(n)$ is the input image block, and d_k is the k^{th} visual word in class vocabulary. After labeling, we can calculate the histogram of the labels of the input image by Eq. (2).

$$h_s^c = \langle h_s^c(1), h_s^c(2), \dots, h_s^c(C_M) \rangle$$

$$h_s^c(k) = \frac{1}{N} \sum_{n=1}^N \delta(L(B_s^c(n)) - k), k = 1 \dots C_M \quad (2)$$

For training images of a class, the total number of partitioned blocks is $S \times N$. We calculate the label histogram for each class, and then sort the usage frequency of visual words. In our work, the top P most frequently used visual words will be selected to form total vocabulary. The construction of total vocabulary not only keeps the characteristics of each class and prevents over-merge but also avoids the complex merging procedure.

According to the characteristics of visual words, we classify the visual words into macro sense or micro sense. The former is with smooth content and the latter contains high activity content such as edges or obvious textures. Fig. 2 is an example to illustrate the concept.

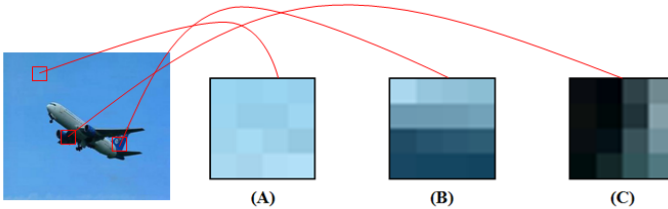


Fig. 2. Sample image, (A) Macro sense, (B) and (C) Micro sense visual words

In our work, the variation of a visual word is used to distinguish macro-sense word from micro-sense word. Let a total visual vocabulary with size T_M be expressed as $\mathbf{D}(k) = \{d_k | k = 1, 2, \dots, T_M\}$, where d_k is a visual word with size of 4×4 ; each component of the vector contains three color channels (d_k^R, d_k^G, d_k^B). For each visual word d_i , we sort the RGB channels respectively in an ascending order, and then calculate the variation ($V_R(d_i), V_G(d_i), V_B(d_i)$) by using Eq. (3),

$$V_c(d_i) = \sqrt{\left(\left[\frac{d_i^{c8} + d_i^{c9}}{2}\right] - \left[\frac{d_i^{c1} + d_i^{c2}}{2}\right]\right)^2 + \left(\left[\frac{d_i^{c15} + d_i^{c16}}{2}\right] - \left[\frac{d_i^{c8} + d_i^{c9}}{2}\right]\right)^2} \quad (3)$$

where superscripts c is the channel, and d_i^{cn} is the n th order element after sorting for each channel. The macro and micro sense visual words are then determined by

$$\begin{aligned} & \text{if } \max(V_R(d_i), V_G(d_i), V_B(d_i)) \leq T \\ & d_i \in D^{MAC}(i) = \{d_i^{Mac}, i = 1, \dots, N_{MAC}\} \\ & \quad \text{else} \\ & d_i \in D^{mic}(i) = \{d_i^{mic}, i = 1, \dots, N_{mic}\}, \end{aligned} \quad (4)$$

where $D^{MAC}(i)$, d_i^{Mac} are macro sense vocabulary and visual word, and $D^{mic}(i)$, d_i^{mic} are micro sense vocabulary and visual word, respectively. $\max(\cdot)$ is a function of selecting maximum element in (\cdot) , and the T is a threshold, which is determined by experiments. However, there are high redundancies in these two types of visual words. Thus, to construct the macro vocabulary and micro vocabulary, further merging of visual words for both types are necessary. To balance the effectiveness and representativeness, the usage frequency and minimum word distance are combined into the merging procedure.

In our work, the sizes of macro-sense vocabulary and micro-sense vocabulary, N_{MAC} and N_{mic} , will be determined by extensive experiments.

3 Image Description and Similarity Measure

For image description, the input block is first categorized as macro-sense or micro-sense using the rule same as Eq. (4). If the input block belongs to macro-sense, it is labeled with macro vocabulary; otherwise with micro sense vocabulary.

The image description is the combination of macro sense histogram and micro sense histogram, which is expressed as $H_s^c(q) = (H_i^{MAC}(q), H_j^{mic}(q))$, where $i=1, \dots, N_{MAC}$, $j=1, \dots, N_{mic}$. Based on the histogram based image description, we can define the similarity of images [2] as Eq. (5) and (6).

$$\begin{aligned} S^{MAC(or mic)}(q, l) &= \sum_{i=1(or j=1)}^{N_{MAC}(or N_{mic})} \left(1 - \left|H_{i(or j)}^{MAC(or mic)}(q) - H_{i(or j)}^{MAC(or mic)}(l)\right|\right) \\ & \quad \times \min\left(H_{i(or j)}^{MAC(or mic)}(q), H_{i(or j)}^{MAC(or mic)}(l)\right) \end{aligned} \quad (5)$$

$$S(q, l) = S^{MAC}(q, l) \times w_1 + S^{mic}(q, l) \times w_2, \quad (6)$$

where $S^{MAC}(q, l)$ is the similarity of image q and l in macro sense histogram, and $S^{mic}(q, l)$ is the micro sense similarity, w_1, w_2 are the weighting value, and $S(q, l)$ is the overall similarity.

In order to evaluate the performance, we also establish a simple categorization model to test the potential of the new visual vocabulary. The categorization scheme does not apply the advance technology; our purpose is to prove the effectiveness of the proposed visual vocabulary. The methodology of categorization is not the main issue in this paper.

For categorization of image, we select the training image from each class of image, and then the class feature model is built accordingly. The training procedure is shown in Fig. 3. We select the number of S training image from each class C , and expressed as $T_s^c = \{t_s^c | c = 1, \dots, C, s = 1, \dots, S\}$, in our work the S is set 5. For simplicity, the description of training images is $H(t_s^c) = (H_C^{MAC}, H_C^{mic})$, where H_C^{MAC} and H_C^{mic} are the macro and micro sense description, respectively. The class feature model is defined as,

$$H_c = \frac{1}{S} \sum_{s=1}^S H(t_s^c) = (H_C^{MAC}, H_C^{mic}) \tag{7}$$

where $H(t_s^c)$ is the histogram of training image t_s^c , H_C is the class feature model, $c=1 \dots C$. When image is input to categorize, the description of input image I is calculated, and then the class feature model is used to determine the image class by Eq (6).

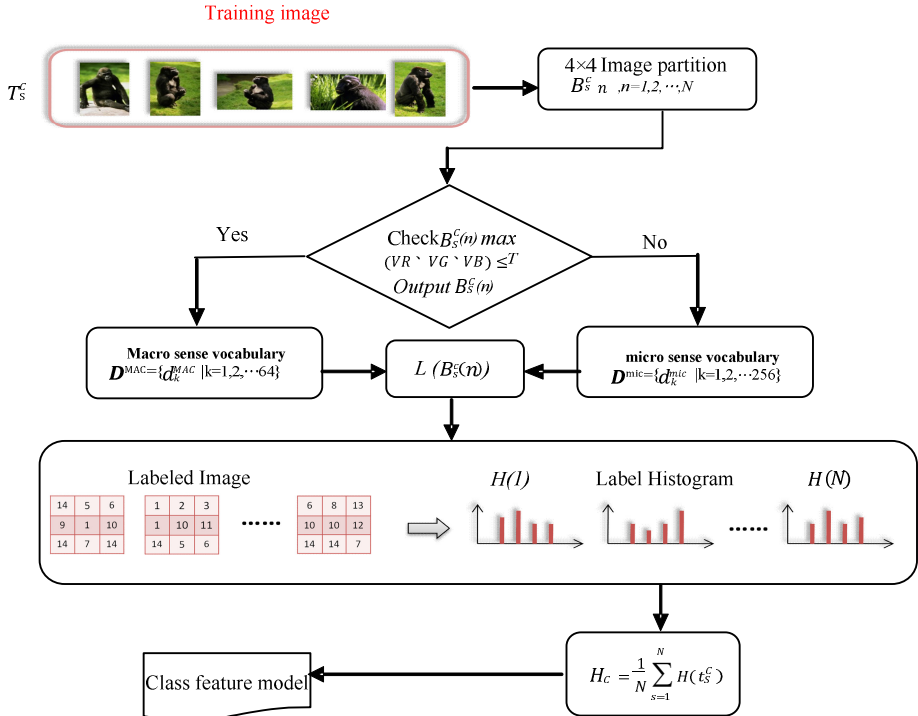


Fig. 3. The flow chart of class feature model construction

$$\text{Cate}(I) = c = \min_c \|H_I - H_c\| \tag{8}$$

where $\text{Cate}(I)$ is the categorize function, H_c is the class feature model, and H_I is the description of input image, $\|H_I - H_c\|$ is the distance of the histogram as in Eq. (5).

4 Experimental Results

We use a database with a variety of images (31 classes, 3901 images) from Corel’s photo to test the performance of the proposed method. To balance the performance and computation, in our work, the vocabulary size is set to 320 with $M:N=64:256$.

We will test the performance of proposed visual vocabulary by simple image categorization. We do not deal with categorization methodology, the purpose is to verify the effectiveness of proposed visual vocabulary. For simplicity, the selected 150 test images that organized as Fig. 4. There are three main categories, i.e., 30 for each category, and each main category has three sub-categories, i.e., 10 for each sub-category. If the categorization is based on main category, the weighting value in similarity measure is set (0.7:0.3), and for sub-category the weighting value is set (0.3:0.7), respectively.








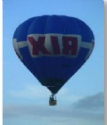
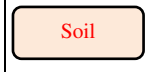





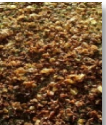

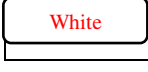



Main categories by global sense	Sub categories					
		Orangutans		Waterfall		Leaf
		Helicopter		Airplane		Hot air balloon
		Elephant		Fighter		Leopard
		Mural		Dried leaf		Sunset
		Dinosaur		Duck		Pot plant

Fig. 4. Global sense category defined in each row and their related sub-category



Fig. 5. The examples of mis-classification

5 Conclusion

In this paper, we have proposed a systematical approach to construct a discriminative visual vocabulary with macro and micro sense of visual words. We also present an effective image description method based on the macro and micro visual vocabulary. In order to evaluate the performance of proposed visual vocabulary, the image categorization is simulated. The experiments indicate the visual vocabulary achieves promising results for retrieval. Therefore, we can conclude that the proposed visual vocabulary can effectively extract the visual features from images.

Acknowledgement. This work was supported by the National Science Council Granted NSC 102-2221-E-214-040-

References

1. Zhu, L., Rao, A., Zhang, A.: Theory of Keyblock-Based Image Retrieval. *ACM Transaction on Information Systems*, 224–257 (2002)
2. Yang, N.C., Chang, W.H., Kuo, C.M., Li, T.H.: A fast MPEG-7 dominant color extraction with new similarity measure for image retrieval. *Journal of Visual Communication and Image Representation* 19, 92–105 (2008)
3. Jiang, Y.G., Yang, J., Ngo, C.W., Hauptmann, A.G.: Representations of Keypoint-Based Semantic Concept Detection: A Comprehensive Study. *IEEE Transactions on Multimedia* 12(1), 42–53 (2010)
4. Li, T., Mei, T., Kweon, I.S., Hua, X.S.: Contextual Bag-of-Words for Visual Categorization. *IEEE Transactions on Circuits And Systems For Video Technology* 21(4), 381–392 (2011)
5. Zhang, S., Tian, Q., Hua, G., Huang, Q., Gao, W.: Generating Descriptive Visual Words and Visual Phrases for Large-Scale Image Applications. *IEEE Transactions on Image Processing* 20(9), 3664–3677 (2011)
6. Kesorn, K., Poslad, S.: An Enhanced Bag-of-Visual Word Vector Space Model to Represent Visual Content in Athletics Images. *IEEE Transactions on Multimedia* 14(1), 211–222 (2012)

7. Perronnin, F.: Universal and Adapted Vocabularies for Generic Visual Categorization. *IEEE Transactions on Pattern Analysis And Machine Intelligence* 30(7), 1243–1256 (2008)
8. Qin, J., Yung, N.C.: Scene categorization via contextual visual words. *Pattern Recognition* 43, 1874–1888 (2010)
9. López-Sastre, R.J., Tuytelaars, T., Rodríguez, F.J.A., Bascón, S.M.: Towards a more discriminative and semantic visual vocabulary. *Computer Vision And Image Understanding* 115, 415–425 (2011)
10. Bolovinou, A., Pratikakis, I., Perantonis, S.: Bag of spatio-visual words for context inference in scene classification. *Pattern Recognition* 46, 1039–1053 (2013)

Part III
**Technologies for Next-Generation
Network Environments**

An Incremental Algorithm for Maintaining the Built FUSP Trees Based on the Pre-large Concepts

Chun-Wei Lin^{1,2}, Wensheng Gan¹, Tzung-Pei Hong^{3,4}, and Raylin Tso⁵

¹Innovative Information Industry Research Center (IIIRC),

²Shenzhen Key Laboratory of Internet Information Collaboration
School of Computer Science and Technology

Harbin Institute of Technology Shenzhen Graduate School
HIT Campus Shenzhen University Town, Xili, Shenzhen 518055 P.R. China

³Department of Computer Science and Information Engineering
National University of Kaohsiung, Kaohsiung, Taiwan, R.O.C.

⁴Department of Computer Science and Engineering
National Sun Yat-sen University, Kaohsiung, Taiwan, R.O.C.

⁵Department of Computer Science
National Chengchi University, Taipei, 11605, Taiwan, R.O.C.

jerrylin@ieee.org, wsgan001@gmail.com, tphong@nuk.edu.tw,
raylin@cs.nccu.edu.tw

Abstract. Mining useful information or knowledge from a very large database to aid managers or decision makers to make appropriate decisions is a critical issue in recent years. In this paper, we adopted the pre-large concepts to the FUSP-tree structure for sequence insertion. A FUSP tree is built in advance to keep the large 1-sequences for later maintenance. The pre-large sequences are also kept to reduce the movement from large to small and vice versa. When the number of inserted sequences is smaller than the safety bound of the pre-large concepts, better results can be obtained by the proposed incremental algorithm for sequence insertion in dynamic databases.

Keywords: Pre-large concept, dynamic databases, sequential pattern mining, sequence insertion, FUSP-tree structure.

1 Introduction

Mining desired knowledge or information to aid managers or decision makers for making the efficient decisions from a very large database is a critical issue in recent years. [1-4, 6, 8, 13]. Among them, sequential patterns mining considers the order sequence data such as Web-click logs, network flow logs or DNA sequences, which is the major issue in real-world applications. For basket analysis, sequential patterns mining can also be used to mine the purchased behaviors of customers to predict whether there is a high probability that when customers buy some products, they will buy some other products in later transactions.

Agrawal et al. first proposed AprioriAll algorithm [3] to level-wisely mine sequential patterns in a batch way. Various algorithms applied in different applications of sequential patterns mining have been proposed to handle the static database [10, 15-17]. Discovered sequential patterns may, however, become invalid since sequences are changed in dynamic databases. Developing an efficient approach to maintain and update the discovered sequential patterns is a critical issue in real-world applications. Lin et al. proposed an incremental FASTUP algorithm [14] to maintain the discovered sequential patterns. Lin et al. designed a fast updated sequential pattern (FUSP)-tree structure and algorithms to handle the sequential patterns in dynamic databases [11-12].

The FASTUP or FUSP-tree algorithms are, however, required to re-scan the original database if the small itemsets or sequences are necessary to be maintained and updated. Hong et al. then extended the pre-large concepts [7] of association-rule mining to level-wisely maintain the sequential patterns in dynamic databases [9]. In this paper, the pre-large concepts are adopted in the FUSP tree to efficiently maintain the discovered sequential patterns for sequence insertion. A FUSP-tree structure is first built to keep only large sequences in the tree, and the pre-large sequences are mined out and kept in a set for later maintenance. The proposed incremental algorithm divides the 1-sequences in the newly inserted sequences into three parts with nine cases. Each case is then performed by the designed algorithm to maintain and update the built FUSP tree. Experimental results also then show that the proposed algorithm has a good performance for incrementally handling new inserted sequences.

2 Review of Related Works

In this section, sequential patterns mining and the pre-large concepts are briefly reviewed.

2.1 Sequential Patterns Mining

In the past, Agrawal et al. designed an AprioriAll algorithm [3] to level-wisely mine sequential patterns in a static database. Lin et al. thus proposed an incremental FASTUP algorithm [14] to maintain sequential patterns in dynamic databases. The FASTUP algorithm is, however, required to re-scan the original database if it is necessary to maintain the discovered sequential pattern, which is large in the added sequences but small in the original database. Hong et al. extended the pre-large concepts of association-rule mining [7] to handle the sequential patterns whether for the sequence insertion [9] or deletion [5]. It is also based on Apriori-like approach [2] to generate-and-test the candidates for deriving the desired sequential patterns. Lin first designed a fast updated sequential pattern (FUSP)-tree and developed the algorithms for efficiently handling sequence insertion in incremental mining [11-12]. Based on the built FUSP tree structure, the discovered sequential patterns can be thus easier maintained.

The FUSP tree [11] is used to store customer sequences with only large 1-sequences in the original database. Based on the FUSP tree, the complete sequential patterns can be derived from it without level-wisely rescanning the original database. An example is given to show the FUSP tree. Assume a database shown in Table 1 is used to build the FUSP tree.

Table 1. An example

CID	Customer Sequence
1	(AC)(F)
2	(CE)(D)(H)
3	(AB)(D)
4	(C)(D)(EF)(H)
5	(AC)(GI)
6	(BC)(DEH)
7	(A)(D)(H)
8	(AF)(DG)
9	(A)(D)(EH)
10	(C)(F)(BD)

Also assume that the upper support threshold is set at 60%, and the lower support threshold is set at 30%. The large 1-sequences are (A), (C), and (D) from which Header_Table can be constructed. The pre-large 1-sequences are then kept in a set of $Pre_Seqs = \{B:3, E:4, F:4, H:5\}$. The built FUSP tree from the database is shown in Figure 1. In Figure 1, only large 1-sequences are kept in the FUSP tree. The link between two connected nodes is marked by the symbol s if the sequence is within the sequence relation in a sequence; otherwise, the link is marked by the symbol i if the sequence is within the itemset relation in a sequence. The built Header_Table is used as an index table to find appropriate items or sequences in the tree. It keeps the large 1-sequences initially in descending order of their counts. Infrequent ones are not used to build the tree. After all customer sequences are processed, the FUSP tree is completely constructed.

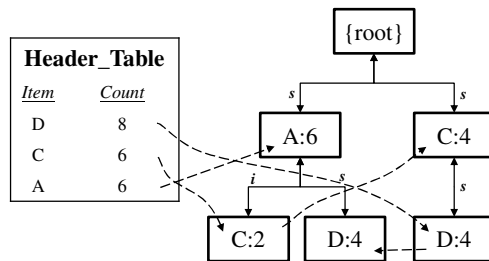


Fig. 1. Initial constructed FUSP tree with its Header_Table

2.2 Pre-large Concepts

A pre-large sequence [9] is not truly large, but has highly probability to be large when the database is updated. A lower support threshold and an upper support threshold are used to respectively define the pre-large and the large concepts. The pre-large concepts act like buffers to reduce the movement of sequences directly from large to small and vice-versa in the maintenance process. Considering an original database and some customer sequences are inserted into the original database, three parts with nine cases in Figure 2 may arise.

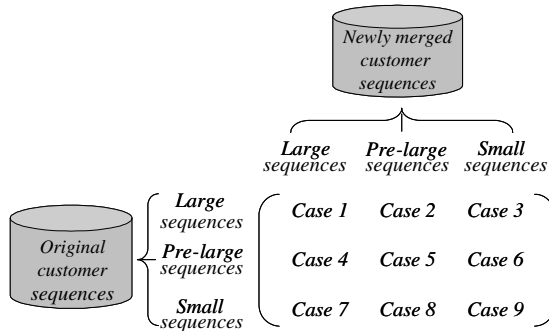


Fig. 2. Nine cases arising from the original database and the inserted sequences

Cases 1, 5, 6, 8 and 9 will not affect the final large sequences. Cases 2 and 3 may remove existing large sequences. Cases 4 and 7 may add new large sequences. If we retain all large and pre-large sequences with their counts in the original database, cases 2, 3 and 4 can be easily handled. In the maintenance phase, the ratio of newly added customer sequences to original customer sequences is usually very small. This is more apparent when the database grows larger. It has been formally shown that the sequences in Case 7 cannot possibly be large in the updated database as long as the number of customer sequences is small compared to the number of customer sequences in the original database. The formula [9] is shown below.

$$t \leq \frac{(S_u - S_l) \times d}{1 - S_u} - \frac{q \times S_u}{1 - S_u},$$

where t is the number of the newly added customer sequences, and q be the number of newly added customer sequences belongs to old customers, S_u is the upper threshold, S_l is the lower threshold, and d is the number of customer sequences in the original database.

3 Proposed an Incremental Algorithm

A fast updated sequential pattern (FUSP)-tree [11] must built in advance to keep the large sequences from the original database before new transactions or sequences

come. The pre-large 1-sequences are also kept in a set for later maintenance process. When the sequences are inserted into the original database, the proposed incremental algorithm is then performed below in details.

Proposed algorithm:

INPUT: An old database consisting of $(d + t)$ sequences, its corresponding Header_Table, a set of *Pre_Seqs* to keep the pre-large 1-sequences, its corresponding FUSP tree, a lower support threshold S_l , an upper support threshold S_u , and a set of t new inserted sequences.

OUTPUT: An updated FUSP tree.

STEP 1: Set $b = b + q$, where q number of newly inserted sequences belonging to old customers in the original database;

STEP 1: Calculate the safety bound f to determine whether the original database is required to be re-scanned of the new inserted transactions by the formula as [9]:

$$t \leq \frac{(S_u - S_l) \times d}{1 - S_u} - \frac{b \times S_u}{1 - S_u}.$$

STEP 2: Scan the new sequences to get all 1-sequences with their frequencies.

STEP 3: Divide the 1-sequences in STEP 2 into three parts with nine cases according to whether they are large (appears in the Header_Table), pre-large (appears in the set of *Pre_Seqs*) or small (not appears in the Header_Table either in the set of *Pre_Seqs*) in the original database.

STEP 4: For each 1-sequence s which is large in the original database, do the following substeps (**Cases 1, 2 and 3**):

Substep 4-1: Set the count $S^U(s)$ of s in the updated database as:

$$S^U(s) = S^D(s) + S^T(s),$$

where $S^D(s)$ is the frequency of s in the Header_Table (original database) and $S^T(s)$ is the frequency of s in the new transactions.

Substep 4-2: If $S_u \leq S^U(s)/(d+c+t-b)$, update the frequency of s in the Header_Table as $S^U(s)$; put s in the set of *Insert_Seqs*, which will be further processed in STEP 8. Otherwise, if $S_l < S^U(s)/(d+c+t-b) \leq S_u$, remove s from the Header_Table; connect the parent node of s to its child nodes directly in the FUSP tree; put s in the set of *Pre_Seqs* with its updated frequency $S^U(s)$. Otherwise, 1-sequence s becomes small after the database is updated; remove s from the Header_Table and connect each parent node of s directly to its child nodes in the FUSP tree.

STEP 5: For each 1-sequence s which is pre-large in the original database, do the following substeps (**Cases 4, 5 and 6**):

Substep 5-1: Set the new count $S^U(s)$ of s in the updated database as:

$$S^U(s) = S^D(s) + S^T(s).$$

Substep 5-2: If $S_u \leq S^U(s)/(d+c+t-b)$, 1-sequence s will be large after the database is updated; remove s from the set of *Pre_Seqs*; put s with its updated frequency in the set of *Branch_Seqs*; put s in the set of *Insert_Seqs*. Otherwise, if $S_l < S^U(s)/(d+c+t-b) \leq S_u$, 1-sequence s still remains pre-large after the database is updated; update s with its new frequency $S^U(s)$ in the set of *Pre_Seqs*. Otherwise, remove 1-sequence s from the set of *Pre_Seqs*.

STEP 6: For each 1-sequence s which is neither large nor pre-large in the original database but large or pre-large in the new transactions (**Cases 7 and 8**), put s in the set of *Rescan_Seqs*, which is used when rescanning the database in STEP 7 is necessary.

STEP 7: If $t + c \leq f - h$ or the set of *Rescan_Seqs* is *null*, then do nothing; Otherwise, do the following substeps for each 1-sequence s in the set of *Rescan_Seqs*:

Substep 7-1: Rescan the original database to decide the original count $S^D(s)$ of s .

Substep 7-2: Set the new count $S^U(s)$ of s in the updated database as:

$$S^U(s) = S^D(s) + S^T(s),$$

Substep 7-3: If $S_u \leq S^U(s)/(d+c+t-b)$, 1-sequence s will become large after the database is updated; put s in both the sets of *Insert_Seqs* and *Branch_Seqs*; insert the items in the *Branch_Seqs* to the end of the Header_Table according to the descending order of their updated frequencies. Otherwise, if $S_l < S^U(s)/(d+c+t-b) \leq S_u$, 1-sequence s will become pre-large after the database is update; put s with its updated frequency in the set of *Pre_Seqs*. Otherwise, do nothing.

Substep 7-4: For each original transaction with a 1-sequence s existing in the *Branch_Seqs*, if s has not been at the corresponding branch of the FUSP tree for the transaction, insert s at the end of the branch and set its count as 1; otherwise, add 1 to the count of the node s .

STEP 8: Insert the sequences in the *Branch_Seqs* to the end of Header_Table according to the descending order of their updated counts. For each original sequence with a sequence s existing in the *Branch_Seqs*, if s has not been at the corresponding branch of the FUSP tree for the processed sequence, insert s to its corresponding position and set its count as 1; otherwise, add 1 to the count of the node s .

STEP 9: For each new transaction with a 1-sequence s existing in the *Insert_Seqs*, if s has not been at the corresponding branch of the FUSP for the new sequence, insert s to its corresponding position and set its count as 1; otherwise, add 1 to the count of the node s .

STEP 10: If $t + c > f - h$, then set $d = d + t + c$ and set $c = 0$; otherwise, set $c = t + c$.

In STEP 7, a corresponding branch is the branch generated from the large 1-sequences in a transaction and corresponding to the order of 1-sequences appeared in the Header_Table. After STEP 10, the final updated FUSP tree is maintained by the proposed algorithm. The new transactions can then be integrated into the original database. Desired sequential patterns can then be found by the FUSP-growth mining algorithm [11].

4 An Example

In this section, an example is given to illustrate the proposed incremental algorithm for maintaining the discovered sequential patterns based on the built FUSP tree [11]. An original database was shown in Table 1, which consists of 10 customer sequences with nine purchased items. In this example, an upper support S_u and the lower support S_l were respectively set at 30% and 60%. The built FUSP tree was shown in Figure 2. Suppose three new customer sequences shown in Table 2 are inserted into the original database. The proposed incremental algorithm is then performed by the designed steps. The global variables c and b are initially set at 0.

Table 2. Three added customer sequences

CID	Customer sequence
5	(CH)(I)
11	(A)(I)(H)
12	(A)(G)(H)

The value of the first term in Formula 1 [9] is calculated as:

$$f = \frac{(S_u - S_l) \times d}{1 - S_u} = \frac{(0.6 - 0.3) \times 10}{1 - 0.6} = 7.5.$$

Since only one customer sequence with $CID = 5$ in Table 2 belongs to old customers in Table 1, q is thus set at 1, and $b = (b + q) (= 0 + 1) (= 1)$. The value of the second term in Formula 1 [9] is calculated as:

$$h = \frac{b \times S_u}{1 - S_u} = \frac{1 \times 0.6}{1 - 0.6} = 1.5.$$

The customer sequences in Table 2 are firstly scanned to get the 1-sequences and their counts. After that, 1-sequences in the added customer sequences are then divided into three parts with nine cases. The designed algorithm is then performed to maintain and update the built FUSP tree. The final updated FUSP tree is then shown in Figure 3.

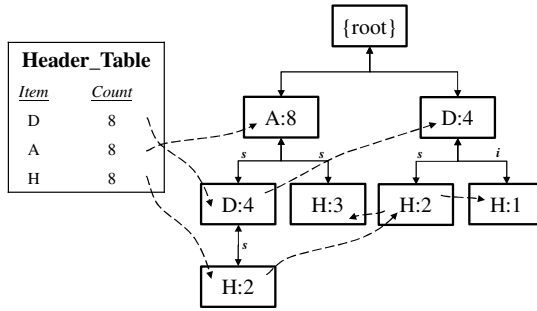


Fig. 3. Initial constructed FUSP tree with its Header_Table

Based on the FUSP tree in Figure 3, the desired large sequences then be found by the FUSP-growth approach [11].

5 Experimental Results

Experiments were made to compare the performance of FUSP-TREE-BATCH algorithm [11], FUSP-TREE-INS algorithm [11], and the proposed incremental algorithm. A real database called BMSWebView-1 [18] is used to evaluate the performance of the proposed incremental algorithm. In the experiments, the execution time and the number of tree nodes are then compared to show the performance of the proposed incremental algorithm at different number of minimum support thresholds. To evaluate the performance of the proposed algorithm at different minimum support thresholds, S_l values are respectively set at S_u values minus 0.21% for BMSWebview-1 database. The results are respectively shown from Figures 4 to 5.

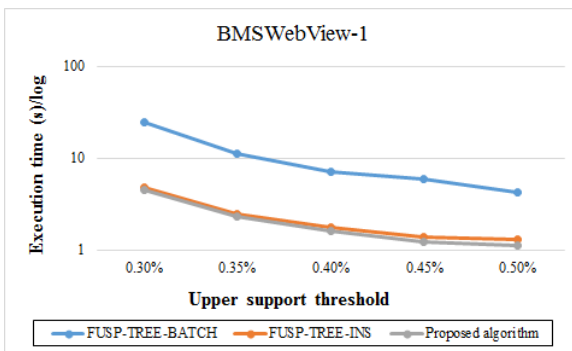


Fig. 4. Comparisons of execution times

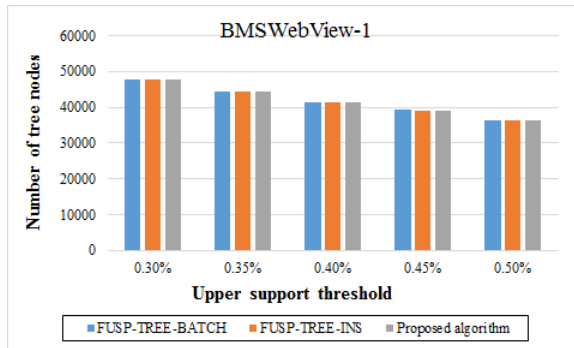


Fig. 5. Comparisons of tree nodes

From Figures 4 and 5, it can be obvious to see that the proposed algorithm runs faster than the FUSP-TREE-BATCH and FUSP-TREE-INS algorithms and generates nearly the same number of tree nodes compared to the other two algorithms. The proposed algorithm can thus be acceptable in terms of execution time and number of tree nodes.

6 Conclusion

In this paper, a pre-large concepts are adopted for efficiently maintaining and updating the built FUSP tree for sequence insertion in dynamic databases. A FUSP-tree structure is used to make the updating process become easier. From the experiments, the proposed incremental algorithm can thus achieve a good trade-off between execution time and tree complexity.

Acknowledgement. This research was partially supported by the Shenzhen Peacock Project, China, under grant KQC201109020055A, by the Natural Scientific Research Innovation Foundation in Harbin Institute of Technology under grant HIT.NSRIF.2014100, by the National Science Council of the Republic of China under Contract no. NSC 101-2628-E-004-001-MY2, and by the Shenzhen Strategic Emerging Industries Program under grant ZDSY20120613125016389.

References

1. Agrawal, R., Imielinski, T., Swami, A.: Database mining: A performance perspective. *IEEE Transactions on Knowledge and Data Engineering* 5, 914–925 (2006)
2. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: *The International Conference on Very Large Data Bases*, pp. 487–499 (1994)
3. Agrawal, R., Srikant, R.: Mining sequential patterns. In: *The International Conference on Data Engineering*, pp. 3–14 (1995)
4. Chen, M.S., Han, J., Philips Yu, S.: Data mining: An overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering* 8, 866–883 (1996)

5. Wang, C.Y., Hong, T.P., Tseng, S.S.: Maintenance of sequential patterns for record deletion. In: IEEE International Conference on Data Mining, pp. 536–541 (2001)
6. Han, J., Pei, J., Yin, Y., Mao, R.: Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Mining and Knowledge Discovery* 8, 53–87 (2004)
7. Hong, T.P., Wang, C.Y., Tao, Y.H.: A new incremental data mining algorithm using pre-large itemsets. *Intelligent Data Analysis* 5, 111–129 (2001)
8. Hong, T.P., Lin, C.W., Wu, Y.L.: Incrementally fast updated frequent pattern trees. *Expert Systems with Applications* 34, 2424–2435 (2008)
9. Hong, T.P., Wang, C.Y., Tseng, S.S.: An incremental mining algorithm for maintaining sequential patterns using pre-large sequences. *Expert Systems with Applications* 38, 7051–7058 (2011)
10. Kim, C., Lim, J.H., Ng, R.T., Shim, K.: Squire: Sequential pattern mining with quantities. *Journal of Systems and Software* 80, 1726–1745 (2007)
11. Lin, C.W., Hong, T.P., Lu, W.H., Lin, W.Y.: An incremental fusp-tree maintenance algorithm. In: The International Conference on Intelligent Systems Design and Applications, pp. 445–449 (2008)
12. Lin, C.W., Hong, T.P., Lu, W.H.: An efficient fusp-tree update algorithm for deleted data in customer sequences. In: International Conference on Innovative Computing, Information and Control, pp. 1491–1494 (2009)
13. Lin, C.W., Hong, T.P.: A survey of fuzzy web mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 3, 190–199 (2013)
14. Lin, M.Y., Lee, S.Y.: Incremental update on sequential patterns in large databases. In: IEEE International Conference on Tools with Artificial Intelligence, pp. 24–31 (1998)
15. Nakagaito, F., Ozaki, T., Ohkawa, T.: Discovery of quantitative sequential patterns from event sequences. In: IEEE International Conference on Data Mining Workshops, pp. 31–36 (2009)
16. Pei, J., Han, J., Mortazavi-Asl, B., Wang, J., Pinto, H., Chen, Q., Dayal, U., Hsu, M.C.: Mining sequential patterns by pattern-growth: The prefixspan approach. *IEEE Transactions on Knowledge and Data Engineering* 16, 1424–1440 (2004)
17. Ren, J.M., Jang, J.R.: Discovering time-constrained sequential patterns for music genre classification. *IEEE Transactions on Audio, Speech, and Language Processing* 20, 1134–1144 (2012)
18. Zheng, Z., Kohavi, R., Mason, L.: Real world performance of association rule algorithms. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 401–406 (2001)

Another Improvement of RAPP: An Ultra-lightweight Authentication Protocol for RFID

Xinying Zheng¹, Chien-Ming Chen^{1,2}, Tsu-Yang Wu^{1,2}, Eric Ke Wang^{1,2},
and Tsui-Ping Chung³

¹ School of Computer Science and Technology, Harbin Institute of Technology
Shenzhen Graduate School, Shenzhen, China

xinying_15@163.com

² Shenzhen Key Laboratory of Internet Information Collaboration, Shenzhen, China
dr.chien-ming.chen@ieee.org, wutsuyang@gmail.com, wk_hit@hitsz.edu.cn

³ Department of Industrial Engineering, Jilin University, Nanling Campus,
Changchun, China
tpchung@jlu.edu.cn

Abstract. RFID technology has received increasing attention; however, most of the RFID products lack security due to the hardware limitation of the low-cost RFID tags. Recently, an ultra-lightweight authentication protocol named RAPP has been proposed. Unfortunately, RAPP is insecure against several attacks. In this paper, we propose an improvement of RAPP. Security analysis demonstrated that our protocol can resist several kinds of attacks.

Keywords: RFID, mutual authentication, security protocol.

1 Introduction

RFID (Radio Frequency IDentification) is a technique for identifying objects via radio frequency. It has received increasing attention in many applications such as supply chain management systems, transportation, access control systems, ticketing systems and animal identification, etc.

An RFID system is composed of three components: a set of tags, RFID readers and one or more backend servers. A backend server stores the related information of tags, calculates the computational processes when authenticates a tag. An RFID reader (called as reader in this paper) accesses a backend server via secure network channel, and then acquires the information related to the tags. RFID tags are small electronic devices which composed of antennas, microprocessors and memory storages. A tag communicates with a reader by using radio frequency signals transmitting from the reader.

Security and privacy issues are concerned mostly in RFID applications. As a result, researchers have proposed many RFID authentication protocols. The RFID authentication protocol can be categorized into 4 classes. The first class

refers to protocols which apply conventional cryptographic functions. The second class refers to protocols that apply random number generator and one-way hash function. The third class refers to protocols that apply random number generator and Cyclic Redundancy Code checksum. The last one refers to those protocols that apply simple bitwise operations (such as XOR, AND, OR, etc.). Generally, the fourth class is treated as ultra-lightweight level.

Several ultra-lightweight authentication protocols for RFID have been proposed. However, most of these protocols are insecure. In this paper, we improved a well-known protocol named RAPP[1]. We also provide a detailed security analysis of the proposed protocol.

2 Related Work

With the rapidly growth of network technology, security issues have been concerned in various network environments [2–10]. In the RFID environment, security issues also receive increasing attention recently. In this paper, we put emphasis on RFID authentication protocol, especially focus on ultra-lightweight authentication protocols.

An ultra-lightweight authentication protocol means that it utilizes only simple bitwise operations on the tags. Several ultra-lightweight protocols have been proposed. Peris-Lopez et al. proposed a family of ultra-lightweight protocols [11–13] in 2006. Later, these protocols are demonstrated to be vulnerable to de-synchronization attacks and full-disclosure attacks [14, 15]. In 2007, Chien proposed another protocol named SASI [16]. However, SASI is vulnerable to de-synchronization attack [17–19], traceability attack [20] and full-disclosure attack [19, 21]. Although Pedro Peris-Lopez et al. [22] attempted to improve SASI, this work [22] is insecure against de-synchronization attack [23–25]. In 2009, Mathieu David et al. [26] introduced another protocol. Unfortunately, this work suffered from a new full-disclosure attack and a traceability attack [27]. In 2011, Aras Eghdamian et al. [28] proposed another ultra-lightweight protocol. However, [29] pointed out this protocol is vulnerable to full-disclosure attacks.

In 2012, Tian et al. [1] proposed a new ultra-lightweight RFID protocol named RAPP. The authors claimed that RAPP can withstand various attacks and provide strong data confidentiality and integrity. Unfortunately, several research have demonstrate that [29–32] is vulnerable various kinds of attacks. Fig. 1 shows the relation of the above protocols.

3 Security Requirement for RFID Authentication Protocol

3.1 Security Requirements

To defend against the common seen threats, the design of RFID systems should satisfy the following security requirements.

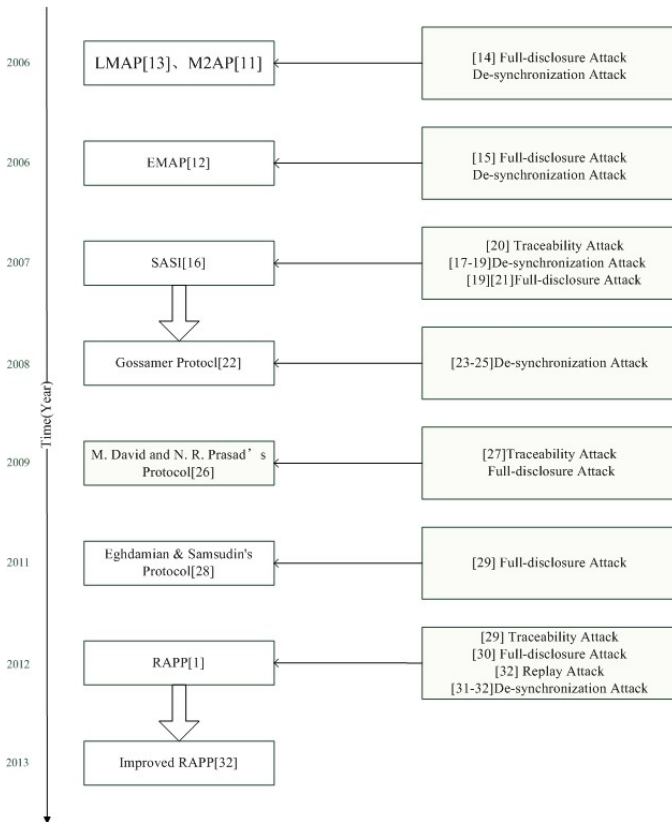


Fig. 1. Related Work of Ultra-Lightweight Authentication Protocol

- Uniqueness: Every tag in an RFID system should be unique, which means an RFID reader should be able to distinguish any RFID tag from the others. This requirement can be satisfied by assigning an unique identifier to each tag, and then the tag responds its identifier to the reader's queries.
- Reader-to-tag authentication: The reader should be able to confirm the identity information that an tag claimed is true. In most of the RFID application, the tag is used to uniquely identify an object or a person. If the reader-to-tag authentication cannot be fulfilled, anyone can forge an RFID tag with another tag's identity information, and then disguise itself as the genuine one.
- Tag-to-reader authentication: Since the adversary may use other invalid readers to query and collect data from the tags without arising carrier's attention. Thus, before transmitting any sensitive data, the tag should verify the reader's identity and authenticate the claimed identity is valid or not.

- Mutual authentication: An RFID system fulfills the property of mutual authentication if it satisfies both reader-to-tag authentication and tag-to-reader authentication.
- Integrity: The message receiver should be able to verify that the received messages has not been modified during transmission. That is, the attacker should not be able to forge a message to substitute the original one.
- Forward secrecy: The session key which derived from a secret key that used in one session will not compromise if the secret key is compromised in the future.
- Anonymity: The information emitted from a tag cannot be link to a product, a person or even the tag itself.
- Resistance to compromising attacks: It is difficult to prevent an adversary from stealing valid tags or readers and then physically compromising them. What we concerned is the impact to the entire system when the adversary acquires the stored data in these compromised devices. With the secrets stored inside the tag, the adversary may figure out a way to forge another valid tag without compromising it. In order to resist to such situation, the secrets should be independent among the tags.
- Resistance to denial-of-service attack: In RFID authentication protocol, the asynchronous data between the reader and the tag will result in authentication failure. And the tag can no longer be scanned by the readers, thus, its service is unavailable. As the result, the protocol should handle these data carefully, and maintain data recovery scheme.

4 The Proposed Protocol

In this section, we describe our protocol which is modified from RAPP [1]. Notations used in this paper are shown in Table 1.

Table 1. Notations

Notation	Description
\oplus	Bitwise XOR operation
$wt(x)$	Hamming weight of the binary string x
$f(x, y)$	A secure lightweight pseudo random function (PRF) which takes two inputs x, y and outputs a pseudo random value where $f(x, y) \neq f(y, x)$.
$Rot(x, y)$	Circular left rotation binary string x by $wt(y)$ bit(s)
$Per(x, y)$	The permutation operation of x according to y
r^i	A random number used for i th authentication
K_1^i, K_2^i	Two keys used for i th authentication
L	The bit length of one pseudonym or one key

Fig. 2 illustrates our design. Each tags has its static identification (ID), and pre-shares a pseudonym (IDS) and two keys with the reader. In our protocol, a reader and a tag authenticate to each other. After that, both reader and tag

update the pseudonym and related keys individually. As shown in Fig. 2, reader not only stores new pseudonym IDS^{new} and key K_1^{new}, K_2^{new} , but also keeps old pseudonym IDS^{old} and key K_1^{old}, K_2^{old} . This is because we try to resist replay attacks and de-synchronization attacks.

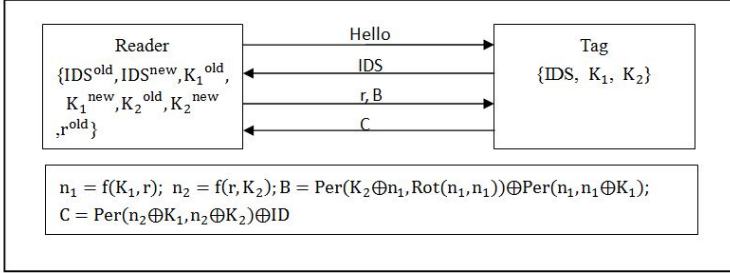


Fig. 2. The Proposed Protocol

As described above, after i th authentication round, the reader keeps $IDS^i, IDS^{i-1}, K_1^i, K_2^i, K_1^{i-1}, K_2^{i-1}, r^i$. The detailed procedures of the $i + 1$ th authentication are listed as follows.

Step 1 A reader sends “Hello” message to a tag to initiate new protocol.

Step 2 This tag responds to the reader with its IDS^i .

Step 3 The reader checks the freshness of this received IDS . If the reader receives IDS^i , it generates a random value r^{i+1} and calculates n_1 and n_2 with K_1^i, K_2^i and r^{i+1} . It then calculates B and transmits B and r^{i+1} to the tag.

After transmitting B and r^{i+1} , the reader updates IDS^{i+1}, K_1^{i+1} and K_2^{i+1} where $IDS^{i+1} = \text{Per}(IDS^i, n_1 \oplus n_2) \oplus K_1^i \oplus K_2^i, K_1^{i+1} = \text{Per}(K_1^i, n_1) \oplus K_2^i$ and $K_2^{i+1} = \text{Per}(K_2^i, n_2) \oplus K_1^i$. The reader also keeps IDS^i, K_1^i, K_2^i and r^{i+1} to prevent replay attacks and de-synchronization attacks.

Step 4 Upon receiving the messages, the tag calculates n_1 and B' with r^{i+1}, K_1^i, K_2^i . If B' equals B , the tag calculates C and sends it to reader.

After sending C to the reader, the tag also calculates IDS^{i+1}, K_1^{i+1} and K_2^{i+1} similar to the equations in Step 3.

Step 5 The reader checks the received C with its secrets.

On the other hand, the reader may receive IDS^{i-1} in step 2. This is because the tag did not update its keys and IDS in the last authentication round for some reasons. As a result, it calculates B with $K_1^{i-1}, K_2^{i-1}, r^i$ and transmits B, r^i to the tag.

5 Security Analysis

In this section, we show that our design is secure against the following attacks.

Replay attack. The replay of tags message will not do harm to our protocol, since the reader stores the random numbers r of the last authentication round. If a reader receives an old *IDS*, indicating that the tag did not get messages r and B in the last round and update its secrets. In this case, the reader uses the old r to continue the protocol.

De-synchronization attack. If an adversary attempts to de-synchronize the shared values between tags and readers in our protocol, he can intercept the message B or message C , or let the reader and the tag use different n_1 and n_2 to update their data. Actually, intercepting C is useless because both the reader and tag have updated their data before C is sent. Besides, intercepting B will cause the reader update its *IDS* and keys, but the tag does not. Once the reader receives the old *IDS*, it will use the old r to continue the protocol. Moreover, letting the reader and the tag use different n_1 and n_2 to update their data is impossible because n_1 and n_2 are calculated using a pseudo random function. That is, an adversary cannot find the direct relationship between r and B or C .

Full-disclosure attack. The main idea behind this attack on RAPP is that the attacker can modify the message A and B to deduce the relationship of adjoining bit of n_1 . However, In this protocol, an adversary cannot obtain K_1 ; thus, n_1 cannot be disclosed.

Traceability attack. An adversary cannot find the Hamming weight of n_1 , n_2 or any other useful values, since all the values will be updated after each protocol round.

6 Comparison

In this section, we compare the performance of our protocol with RAPP in terms of computation operation, the storage requirement and the communication cost. As shown in Table 2, our scheme has a better performance. Note that L means the bit length of one pseudonym or one key.

Table 2. Notations

	RAPP	Our Protocol
Computation	$17 \oplus$, 11 permutations, 2 rotations	$11 \oplus$, 6 permutations, 1 rotations
Storage requirement	5L	4L
Communication	7L	5L

7 Conclusion

In this paper, we propose an improvement of RAPP. Security analysis demonstrated that our protocol can resist several kinds of attacks. We also show that our protocol has better performance than RAPP.

Acknowledgement. The work of Chien-Ming Chen was supported in part by the Project HIT.NSRIF. 2014098 Supported by Natural Scientific Research Innovation Foundation in Harbin Institute of Technology, in part by Shenzhen Peacock Project, China, under Contract KQC201109020055A, and in part by Shenzhen Strategic Emerging Industries Program under Grant ZDSY20120613125016389. the work of Eric Ke Wang was supported in part by National Natural Science Foundation of China (No.61100192), and Shenzhen Strategic Emerging Industries Program under Grants No. JCYJ20120613151032592.

References

1. Tian, Y., Chen, G., Li, J.: A new ultralightweight rfid authentication protocol with permutation. *IEEE Communications Letters* 16(5), 702–705 (2012)
2. Chen, C.M., Lin, Y.H., Chen, Y.H., Sun, H.M.: Sashimi: secure aggregation via successively hierarchical inspecting of message integrity on wsn. *Journal of Information Hiding and Multimedia Signal Processing* 4(1), 57–72 (2013)
3. Wei-Chi, K., Chien-Ming, C., Hui-Lung, L.: Cryptanalysis of a variant of peyравian-zunic’s password authentication scheme. *IEICE Transactions on Communications* 86(5), 1682–1684 (2003)
4. Wu, T.Y., Tseng, Y.M.: Further analysis of pairing-based traitor tracing schemes for broadcast encryption. *Security and Communication Networks* 6(1), 28–32 (2013)
5. Chen, C.M., Wang, K.H., Wu, T.Y., Pan, J.S., Sun, H.M.: A scalable transitive human-verifiable authentication protocol for mobile devices. *IEEE Transactions on Information Forensics and Security* 8(8), 1318–1330 (2013)
6. Hong, T.P., Lin, C.W., Yang, K.T., Wang, S.L.: Using tf-idf to hide sensitive itemsets. *Applied Intelligence*, 1–9 (2013)
7. Chien-Ming, C., Wei-Chi, K.: Stolen-verifier attack on two new strong-password authentication protocols. *IEICE Transactions on Communications* 85(11), 2519–2521 (2002)
8. Wu, T.Y., Tseng, Y.M.: Publicly verifiable multi-secret sharing scheme from bilinear pairings. *IET Information Security* 7(3), 239–246 (2013)
9. Chen, C.M., Chen, Y.H., Lin, Y.H., Sun, H.M.: Eliminating rouge femtocells based on distance bounding protocol and geographic information. *Expert Systems with Applications* 41(2), 426–433 (2014)
10. Sun, H.M., Wang, H., Wang, K.H., Chen, C.M.: A native apis protection mechanism in the kernel mode against malicious code. *IEEE Transactions on Computers* 60(6), 813–823 (2011)
11. Peris-Lopez, P., Hernandez-Castro, J.C., Estevez-Tapiador, J.M., Ribagorda, A.: M²AP: A Minimalist Mutual-Authentication Protocol for Low-cost RFID Tags. In: Ma, J., Jin, H., Yang, L.T., Tsai, J.J.-P. (eds.) *UIC 2006*. LNCS, vol. 4159, pp. 912–923. Springer, Heidelberg (2006)
12. Peris-Lopez, P., Hernandez-Castro, J.C., Estevez-Tapiador, J.M., Ribagorda, A.: EMAP: An efficient mutual-authentication protocol for low-cost RFID tags. In: Meersman, R., Tari, Z., Herrero, P. (eds.) *OTM 2006 Workshops*. LNCS, vol. 4277, pp. 352–361. Springer, Heidelberg (2006)

13. Peris-Lopez, P., Hernandez-Castro, J., Estevez-Tapiador, J., Ribagorda, A.: LMAP: A Real Lightweight Mutual Authentication Protocol for Low-Cost RFID tags. In: Proc. of the 2nd Workshop on RFID Security (2006)
14. Li, T., Wang, G.: Security Analysis of Two Ultra-Lightweight RFID Authentication Protocols. In: Venter, H., Eloff, M., Labuschagne, L., Eloff, J., von Solms, R. (eds.) *New Approaches for Security, Privacy and Trust in Complex Environments*. IFIP, vol. 232, pp. 109–120. Springer, Boston (2007)
15. Li, T., Deng, R.: Vulnerability Analysis of EMAP—an Efficient RFID Mutual Authentication Protocol. In: Proc. of the 2nd Inter. Conf. on Availability, Reliability and Security, pp. 238–245 (2007)
16. Chien, H.Y.: SASI: A New Ultralightweight RFID Authentication Protocol Providing Strong Authentication and Strong Integrity. *IEEE Trans. on Dependable and Secure Computing* 4(4), 337–340 (2007)
17. Cao, T., Bertino, E., Lei, H.: Security analysis of the sasi protocol. *IEEE Transactions on Dependable and Secure Computing* 6(1), 73–77 (2009)
18. Sun, H.M., Ting, W.C., Wang, K.H.: On the Security of Chien’s Ultralightweight RFID Authentication Protocol. *IEEE Trans. on Dependable and Secure Computing* 8(2), 315–317 (2009)
19. D’Arco, P., De Santis, A.: On ultralightweight rfid authentication protocols. *IEEE Transactions on Dependable and Secure Computing* 8(4), 548–563 (2011)
20. Phan, R.W.: Cryptanalysis of a New Ultralightweight RFID Authentication Protocol – SASI. *IEEE Trans. on Dependable and Secure Computing* 6(4), 316–320 (2009)
21. Hernandez-Castro, J.C., Tapiador, J.M., Peris-Lopez, P., Quisquater, J.J.: Cryptanalysis of the sasi ultralightweight rfid authentication protocol with modular rotations. arXiv preprint arXiv:0811.4257 (2008)
22. Peris-Lopez, P., Hernandez-Castro, J.C., Tapiador, J.M.E., Ribagorda, A.: Advances in ultralightweight cryptography for low-cost RFID tags: Gossamer protocol. In: Chung, K.-I., Sohn, K., Yung, M. (eds.) *WISA 2008*. LNCS, vol. 5379, pp. 56–68. Springer, Heidelberg (2009)
23. Bilal, Z., Masood, A., Kausar, F.: Security analysis of ultra-lightweight cryptographic protocol for low-cost rfid tags: Gossamer protocol. In: *International Conference on Network-Based Information Systems, NBIS 2009*, pp. 260–267. IEEE (2009)
24. Yeh, K.H., Lo, N.: Improvement of two lightweight rfid authentication protocols. *Information Assurance and Security Letters* 1, 6–11 (2010)
25. Tagra, D., Rahman, M., Sampalli, S.: Technique for preventing dos attacks on rfid systems. In: *2010 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pp. 6–10. IEEE (2010)
26. David, M., Prasad, N.R.: Providing strong security and high privacy in low-cost rfid networks. In: Schmidt, A.U., Lian, S. (eds.) *MobiSec 2009*. LNICST, vol. 17, pp. 172–179. Springer, Heidelberg (2009)
27. Hernandez-Castro, J.C., Peris-Lopez, P., Phan, R.C.W., Tapiador, J.M.: Cryptanalysis of the david-prasad rfid ultralightweight authentication protocol. In: *Radio Frequency Identification: Security and Privacy Issues*. Springer (2010) 22–34
28. Eghdamian, A., Samsudin, A.: A secure protocol for ultralightweight radio frequency identification (rfid) tags. In: Abd Manaf, A., Zeki, A., Zamani, M., Chuprat, S., El-Qawasmeh, E. (eds.) *ICIEIS 2011, Part I*. CCIS, vol. 251, pp. 200–213. Springer, Heidelberg (2011)

29. Avoine, G., Carpent, X.: Yet another ultralightweight authentication protocol that is broken. In: Hoepman, J.-H., Verbauwhede, I. (eds.) RFIDSec 2012. LNCS, vol. 7739, pp. 20–30. Springer, Heidelberg (2013)
30. Shao-hui, W., Zhijie, H., Sujuan, L., Dan-wei, C.: Security analysis of rapp an rfid authentication protocol based on permutation. Technical report, Cryptology ePrint Archive, Report 2012/327 (2012)
31. Ahmadian, Z., Salmasizadeh, M., Aref, M.R.: Desynchronization attack on rapp ultralightweight authentication protocol. *Information Processing Letters* 113(7), 205–209 (2013)
32. Zhuang, X., Wang, Z.H., Chang, C.C., Zhu, Y.: Security analysis of a new ultralightweight rfid protocol and its improvement. *Journal of Information Hiding and Multimedia Signal Processing* 4(3) (2013)

A Security System Based on Door Movement Detecting

Ci-Rong Li¹, Chie-Yang Kuan²,
Bing-Zhe He², Wu-En Wu³, Chi-Yao Weng², and Hung-Min Sun²

¹ Department of Management, Fuqing Branch of Fujian Normal University,
Fujian, China

cirongli@gmail.com

² Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan
{cykuan, ckshjerho}@is.cs.nthu.edu.tw, {mnesia1, chiyao.weng}@gmail.com,
hmsun@cs.nthu.edu.tw

³ Department of Mathematics, Soochow University, Taipei, Taiwan
mn@scu.edu.tw

Abstract. Recently, the smartphone devices have become the one of the most popular production in the whole world and the price of smartphone have become cheaper and cheaper. Especially, most of the smartphones have embedded lots of sensors such as light sensor, orientation sensor, accelerometer sensor, etc. Thus, our research is implemented a security application called DoorPass which based on smartphone device sensors. By placing the smartphone behind the door, DoorPass can detect the door movement, and provide some protection to the user. We provided three kinds of notification which are sending sms, make a phone call, and send email. Besides, we also provide three different functions for protection which are track phone, video record, and face detection. By implementing on the smartphone, DoorPass not only can provide the protection but also lower down the cost fee of buying the security hardware and provided convenient, simple and security functions.

Keywords: Security System, Object Moving Detection, and Smart Phone App.

1 Introduction

Nowadays, the requirement of security system (anti-theft system) [1] have been grown up rapidly since people are very concerned the security of their house. Generally, most of the securities devices provided from the security companies are equipped with various embedded motion sensors and camera. By using the sensors, the security devices can provide some detection and do some protection when detecting the abnormal situation. Although, security system is very powerful for providing protection to the user's house, but the cost of the security devices are very expensive and most of the users can't effort it. Thus, how to lower the price and provide the same quality of protection has become an important issue.

In recent years, smartphones are embedded lots of sensors such as gyroscope sensor, orientation sensor[2][3]. They can provide very useful like context awareness[4], and motion-based commands(shake the phone to turn on camera)[5]. Since the use of motion sensors in smartphone has become more widespread, we propose a solution called DoorPass to solve the problems which have mention above. DoorPass is a security application which employ smartphone sensor for detecting the door movement. It can send the notification and provide some protection if detect the door has been opened. By implementing DoorPass in smartphone, it can lower down the cost fee of security devices and provide the security, simple, and lightweight of protection.

2 Related Work

With the rapidly growth of network technology, security issues have been concerned in various network environments [6–13] In this section, we would like to introduce some smartphone sensors which have been used to detect context aware and had been proposed in recent years. Moreover, we will indicate the smartphone sensors which has its advantage and disadvantage because in some situation smartphone sensors may leak users privacy and without users any intentions. Therefore, we will introduce as follows.

- Activity Recognition (Advantage)

In the last few years, activity recognition has been addressed by several works using accelerometer wearable devices [14] [15] [16] [17] [18]. Most of them use the same methodology such as collect a lots of accelerometer data from very large experiment, and then extract the specific feature to analyst the raw data and use some learning machine to classifier what kind of activity are doing from the users now.

- Anti-blurred Images (Advantage)

Most of the time when we try to take some photos from our smartphone camera we may get some blurred images even if the smartphone have some anti-shake functions. In the paper [19] , the author using accelerometer and orientation sensors detect the vibration and implement a method to decrease the blurred images capture and can get the better images quality even the hands have some shake.

- Keystroke Inference Attack (Disadvantage)

In some cases, using smartphone sensors may trade user's privacy without any intension for the user. There exist some cases [20] [21] [22] [23] that based on accelerometer or orientation sensors to infer keystrokes on touch screen. There are four steps need to be complete to inferring a keystroke. First, sensor need to do the sniffer work, then do the preprocessing to extract the feature and training the data get from the sensors. Finally, the dispatcher will return the inference result to the attacker.

- Location Inference Attack (Disadvantage)

In the paper [24] , the authors show that using accelerometer and locate a user position within 200 meters radius of the true location. Since using

sensors in Android don't need any permission declared, it's very easy to let attacker using this kinds of side channel attacks.

The sensors can be used in various way and provide some specific functions, it totally depends on the developers. Even if the sensors may leak the privacy and security caused of malicious developers, we believe there are still lots of good developers and various functions can be done by using the smartphone's sensors.

3 System Design

Since our goal of this research is provided a security application through smartphone which can detect door movement and send notification to the users. Therefore, in this section we would like to propose our solution called DoorPass. There are two parts to introduce our system design. The first part we will point out some requirements of our application. We will explain our application overview on the second part.

3.1 Overview of DoorPass

DoorPass is an intrusion detection application which employs accelerometer and orientation sensors to detect door movement and determine that is unknown people come to user's home. We've provided three different ways to notify the user when detect the door movement and we also have provided four different functions to take action when detect unknown people broke into the house. First, we will describe the notifications and later we will also describe the functions as follow.

Notifications

- SMS

When the user chooses this notification, the user must enter the phone number, then DoorPass will send a message via sms to the user when detect the door have been opened.

- Phone Call

When the user chooses this notification, the user must enter the phone number, then DoorPass will make a phone call to the user when detect the door have been opened.

- Gmail

When the user chooses this notification, the user must enter the Gmail address and password, then DoorPass will send an email via internet to the user when detect the door have been opened.

These three notifications can be used interactively. The user can choose all or none; it's totally relying on the user demand. If the user smartphone have a sim card, then he/she can use the three different notifications. Otherwise, the user only can choose the email option.

Functions

– Track Mode

When the user chooses this function, there are some settings need to be accomplish. First, the user needs to input the phone number Gmail address and Gmail password. Then, it will start the service when DoorPass have detected the smartphone have been taken. DoorPass will send the longitudes and latitudes of the smartphone location every minute via sms or Gmail according to the user settings. DoorPass will try to detect the GPS or Wifi whether which one have been turned on, because these two functions can provide the location information to the DoorPass. By using the information which provided from GPS or Wifi, the user can easily track his/her smartphone via DoorPass. This function will run on the background service and don't have any GUI (graphical user interface), therefore it's hard to find out.

– Record video

When the user chooses this function, there are some settings need to be accomplish. First, the user need to enter the FTP IP (or domain name), username and password of the FTP, and time length of each videos. DoorPass will start record the video when detect the door have been opened and stopped each time length of the videos according to the user settings. On the same time, DoorPass will upload the previous video which have been captured and continue record the next video. For each videos have been uploaded successfully, the videos which stored in smartphone will be deleted. Therefore, the capacity of the smartphone will not be filling full after video have been recorded for a long time.

– Face Detection

Since, the unknown people would break into user's house from the windows or different doors. Therefore, this function can enhance the protection even the unknown people used another ways to break into the house. When DoorPass have detected face, it will notify the user and take action according to the user settings.

– Sound Detection

When the user chooses this function, there are some settings need to be accomplish. First, the user needs to input the decibel, occur times, and minutes. Then, DoorPass will start to detect the voice via the microphone and the interface is same as figure[]. According to the user settings, which mean if DoorPass has received over 5 times greater than 55db value within one minute, it will notify the user according to the user settings. Generally, the decibel inside the room is around 30db~40db, if DoorPass have detected the sound is greater than 60db (recommend), we consider that someone is inside the room.

These functions and notification can be used interactively. Record video can provide the image and would be the important evidence if have captured the face of unknown people who broke into user's house. If someone breaks into user's house by different door or windows, face detection and sound detection can send the notification to the user. The track mode will send the

current location of the smartphone when the smartphone have been taken. Therefore, these four functions are totally relying on the user demand.

4 Implementation

In this section, we will introduce how we implement DoorPass on the Android smartphones. First, we would explain the pattern of sensors reading. Second, we would introduce the Android's component which used on the DoorPass. Finally, we would introduce the processes of using DoorPass.

4.1 Observed Pattern of Sensors Reading

In our application, our goal is trying to detect the door movement by using the orientation sensor and using accelerometer to detect the smartphone whether have been taken.

Orientation Sensor. To detect the door movement, we utilize the unique change pattern of direction on the smartphone during door movement. We've placed the phone behind the door and also place a sticker on the floor to observe the door movement. We've tried to open the door until the door reach the white sticker (shown in figure 1). Since the gap is small enough and it's impossible to fit a person break inside the room, we've adopted 10 degree is an act of door open.

To measure the change of the direction, DoorPass will get an initial value (V_1) from the orientation sensor when the DoorPass have been started. Later, we plus

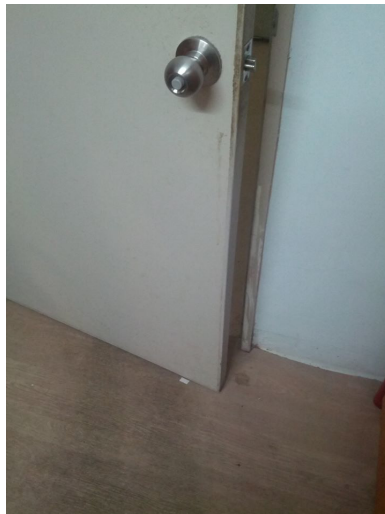


Fig. 1. White sticker on the floor

10 degree to the initial value as V_2 and keep reading the sensor data. If the sensor reading is larger than V_2 , then we consider it's an open door action. In figure 2, we show the orientation sensor readings when smartphone is placing behind the door and is performing different activities. As shown in this comparison, when the door is moving, the fluctuation of wave is very obviously.

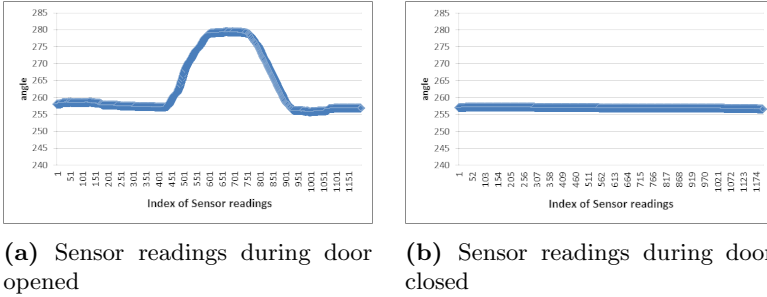


Fig. 2. Sensor readings of orientation sensor

Accelerometer Sensor. To detect the smartphone have been taken event, we utilize the unique change pattern of external force on the smartphone. To measure the change of the external Force F , we used our algorithm $Tsum$ (i.e., $Tsum = |Ax| + |Ay| + |Az|$) to recognize the smartphone pick up event. DoorPass will get an initial value (V_1) from the accelerometer sensor when the DoorPass have been started. Later, we plus 10 degree to the initial value as V_2 and keep reading the sensor data. If the sensor reading is larger than V_2 , then we consider it's a smartphone pick up event. In figure 3, we show the accelerometer sensor readings when smartphone is placing behind the door and is performing different activities. As shown in this comparison, the accelerometer value is always equal to the value of gravity ($9.8ms^{-2}$) when is placing behind the door, and the value change obviously when pick up the smartphone.

4.2 The Process of Using DoorPass

The following step would guide the users for Using DoorPass:

- Step1. Installation

On the beginning, user has to install the DoorPass on his/her Android smartphone. The installation is same as the normal application which can let user download from the Google Play. The installation would be done after the user has confirmed the permission which list on the DoorPass.

- Step2. Set additional functions

After finish the installation, user can turn on the DoorPass and select the notifications and functions he/she needed then input the execution time (hours) that user wants to keep DoorPass running. User can click the set

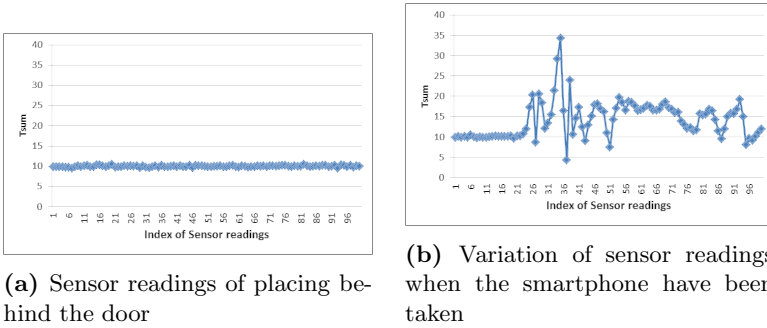


Fig. 3. Sensor readings of accelerometer sensor

button to select the start time for DoorPass when finish the selection of the functions and notifications.

- Step3. Place the smartphone behind the door
User has to place his/her smartphone behind the door before the DoorPass started. If DoorPass detect any door movement, it will do the actions according to user settings. Otherwise, it will finish itself after the execution time over.

As mention above, the setting of DoorPass is very easy and convenient. User just needs a few steps to complete the setting and let DoorPass do the detect jobs.

5 Conclusion

Recently, the smartphone device has become very popular since iOS and Android shown up in four years ago. People use smartphone to access the internet, play games, send email, etc. Moreover, most of the smartphone devices have embedded lots of sensors to increase the user experiences i.e., using temperature sensor to monitor the temperature, using orientation sensor to determine device position and etc. Consequently, there are lots of companies which have provided security system such as Bosch, ADT, Uniforce, and more. Although, these companies can provide very powerful protection to the user's house, but the cost of equipment and installation fee are too expensive and most of the users can't effort this payment. In order to lower the cost of security system and provide protection to the user's home, we propose DoorPass.

DoorPass is a security application which detects the door movement by using the orientation sensor and accelerometer sensor of smartphone. DoorPass also provides some functions such as video record, face detection, and track mode which have been mention in system design section. Furthermore, DoorPass also can notify the user when detect the door have been opened by sending sms, make a phone call, and send email. Therefore, DoorPass can provide the protection to the user's house with a lower cost by using smartphone sensors.

Acknowledgement. This work was supported in part by the National Science Council, Taiwan, under Contracts NSC 100-2628-E-007-018-MY3.

References

1. Security company, http://www.boschsecurity.com.tw/content/language1/html/55_CHT_XHTML.asp
2. Android sensor, <http://developer.android.com/reference/android/hardware/Sensor.html>
3. Comparison of smartphones, http://en.wikipedia.org/wiki/Comparison_of_smartphones
4. Ravindranath, L., Newport, C., Balakrishnan, H., Madden, S.: Improving wireless network performance using sensor hints. In: Proceedings of the 8th USENIX Conference on Networked Systems Design and Implementation, p. 21. USENIX Association (2011)
5. Motorola motox, http://www.motorola.com/us/consumers/Moto-X/moto-x,en_US,pg.html
6. Wang, E.K., Ye, Y., Xu, X.: Location-based distributed group key agreement scheme for vehicular ad hoc network. *International Journal of Distributed Sensor Networks* 2014 (2014)
7. Wei-Chi, K., Chien-Ming, C., Hui-Lung, L.: Cryptanalysis of a variant of peyavian-zunic's password authentication scheme. *IEICE Transactions on Communications* 86(5), 1682–1684 (2003)
8. Chen, C.M., Chen, Y.H., Lin, Y.H., Sun, H.M.: Eliminating rouge femtocells based on distance bounding protocol and geographic information. *Expert Systems with Applications* 41(2), 426–433 (2014)
9. He, B.Z., Chen, C.M., Su, Y.P., Sun, H.M.: A defence scheme against identity theft attack based on multiple social networks. *Expert Systems with Applications* 41(5), 2345–2352 (2014)
10. Sun, H.M., Wang, H., Wang, K.H., Chen, C.M.: A native apis protection mechanism in the kernel mode against malicious code. *IEEE Transactions on Computers* 60(6), 813–823 (2011)
11. Wu, T.Y., Tseng, Y.M.: Further analysis of pairing-based traitor tracing schemes for broadcast encryption. *Security and Communication Networks* 6(1), 28–32 (2013)
12. Chen, C.M., Wang, K.H., Wu, T.Y., Pan, J.S., Sun, H.M.: A scalable transitive human-verifiable authentication protocol for mobile devices. *IEEE Transactions on Information Forensics and Security* 8(8), 1318–1330 (2013)
13. Chen, C.M., Lin, Y.H., Chen, Y.H., Sun, H.M.: Sashimi: secure aggregation via successively hierarchical inspecting of message integrity on wsn. *Journal of Information Hiding and Multimedia Signal Processing* 4(1), 57–72 (2013)
14. He, Z.Y., Jin, L.W.: Activity recognition from acceleration data using ar model representation and svm. In: 2008 International Conference on Machine Learning and Cybernetics, vol. 4, pp. 2245–2250. IEEE (2008)
15. He, Z., Jin, L.: Activity recognition from acceleration data based on discrete cosine transform and svm. In: IEEE International Conference on Systems, Man and Cybernetics, SMC 2009, pp. 5041–5044. IEEE (2009)
16. Bao, L., Intille, S.S.: Activity recognition from user-annotated acceleration data. In: Ferscha, A., Mattern, F. (eds.) PERVASIVE 2004. LNCS, vol. 3001, pp. 1–17. Springer, Heidelberg (2004)

17. Casale, P., Pujol, O., Radeva, P.: Human activity recognition from accelerometer data using a wearable device. In: Vitrià, J., Sanches, J.M., Hernández, M. (eds.) *IbPRIA 2011*. LNCS, vol. 6669, pp. 289–296. Springer, Heidelberg (2011)
18. Ravi, N., Dandekar, N., Mysore, P., Littman, M.L.: Activity recognition from accelerometer data. In: *AAAI*, pp. 1541–1546 (2005)
19. Shin, S.H., Yeo, J.Y., Ji, S.H., Jeong, G.M.: An analysis of vibration sensors for smartphone applications using camera. In: *2011 International Conference on ICT Convergence (ICTC)*, pp. 772–773. IEEE (2011)
20. Xu, Z., Bai, K., Zhu, S.: Taplogger: Inferring user inputs on smartphone touchscreens using on-board motion sensors. In: *Proceedings of the Fifth ACM Conference on Security and Privacy in Wireless and Mobile Networks*, pp. 113–124. ACM (2012)
21. Cai, L., Chen, H.: Touchlogger: inferring keystrokes on touch screen from smartphone motion. In: *Proceedings of the 6th USENIX Conference on Hot Topics in Security*, p. 9. USENIX Association (2011)
22. Owusu, E., Han, J., Das, S., Perrig, A., Zhang, J.: Accessory: password inference using accelerometers on smartphones. In: *Proceedings of the Twelfth Workshop on Mobile Computing Systems & Applications*, p. 9. ACM (2012)
23. Cai, L., Chen, H.: On the practicality of motion based keystroke inference attack. In: Katzenbeisser, S., Weippl, E., Camp, L.J., Volkamer, M., Reiter, M., Zhang, X. (eds.) *Trust 2012*. LNCS, vol. 7344, pp. 273–290. Springer, Heidelberg (2012)
24. Han, J., Owusu, E., Nguyen, L.T., Perrig, A., Zhang, J.: Accomplice: Location inference using accelerometers on smartphones. In: *2012 Fourth International Conference on Communication Systems and Networks (COMSNETS)*, pp. 1–9. IEEE (2012)

Network Performance QoS Prediction

Jaroslav Frnda, Miroslav Voznak, and Lukas Sevcik

Department of Telecommunications, VSB – Technical University of Ostrava,
17. listopadu 15, 70833 Ostrava, Czech Republic
{jaroslav.frnda,miroslav.voznak,lukas.sevcik.st1}@vsb.cz

Abstract. This paper deals with QoS prediction of triple play services in IP networks. Based on our proposed model, speech or video quality can be calculated with regard to policies applied for packet processing by routers and to the level of total network utilization. This new simulating model was implemented in SW tool which enables networkers to predict objective QoS parameters of triple play services and to help them in network design. The contribution of this paper lies in designing a new model capable of predicting the quality of Triple-play services in networks based on IP.

Keywords: Delay, E-Model, Network Performance Monitoring, Packet Loss, PSNR, QoS, SSIM, Triple Play.

1 Introduction

Network convergence that took place during the early 90's of the last century, as well as appearance of NGN (Next Generation Network) concept, allowed the transfer of formerly separate services (voice, video and data) by one common network infrastructure. However, this transition had to deal with some difficulties, as packet networks based on IP protocol had not been designed to transfer delay-sensitive traffic. The difficulties appeared especially at the transfer of voice because, without any supplementary mechanisms securing the quality of service, such a transfer was not capable of providing a high-quality interactive communication similar to standard PSTN. Constant network monitoring, along with network performance intervening as needed, seems to be a method for securing at least minimal QoS level in packet network. Therefore, the purpose of the model described in this paper is to provide a simple monitoring tool capable of predicting qualitative QoS parameters according to network status. The application aims to be an alternative to expensive monitoring tools, as well as a helpful tool for designing network infrastructure with regard to securing at least minimal QoS level.

2 State of the Art

Recently growing interest in voice and video transfer through packet networks based on IP protocol caused that analyses of these services and their behavior in such

networks became more intensive. Logically, the highest emphasis is being put on the transfer of voice, since this service is the most sensitive to an overall network status. References [1, 2, 3] focus on degradation of voice service caused by delay and packet loss. These works use a simplified version of calculation model based on recommendation ITU-T G.107 (also known as E-model) [4] to evaluate the quality of speech, adjusting the model to be suitable especially for packet networks. At the same time, results are used to compare the application results to real practical experiments. Since the final delay and packet loss are factors depending on full network utilization and QoS policy applied to prioritized data flow processing by routers, it is necessary to consider this link as well. Works [5, 6] and [10] analyse in detail the impact of network utilization and set policies on variable component of total delay. In [7] especially the impact of the buffer size for Jitter and packet loss in network is being studied. The analysis of video quality focuses on resistance of video codecs towards packet loss in network which cause the artefacts in video [8, 9]. The measurements in these works show dependence of resolution and bitrate on decrease of video quality through increasing packet loss rate in network – the higher the bitrate and resolution are, the better their resistance towards unwanted effects is. What is still missing, though, is the application of those experiments results to real use. Therefore, this paper attempts to bring a tool which, according to the application of mathematical models based on actual results, would be capable of providing with reliable information on Triple play services quality. Nowadays, the inclination to NGN concept is huge, that is why monitoring of efficiency and evaluation of impacts on QoS is highly important and necessary in order to secure the competitiveness of any multimedia services provider.

3 Methodology

This paper follows directly our previous experiments studying the impact of full network utilization and performance of data prioritization on the final quality of service [5]. For further upgrade of the model, it was essential to analyse the link between different level of network utilization and Triple play services quality (mainly the quality of voice and video). In order to generate VoIP calls (5 for each test and every test has been performed 10 times), the tool *IxChariot* from Ixia company was used. The video was streamed by *VLC player*. Linux tool *iperf* was used for utilization of the network to specific levels. *MSU VQMT* tool served to compare a streamed video with the original one, so that the impact of network utilization could be analysed. Implementation of the computational model was carried out in programming language C#.

3.1 Measured Parameters

E-model, defined by recommendation G.107 by ITU-T [4], was used as an objective evaluation method for voice service. This model is based on calculation of R-factor, as follows:

$$R = R_O - I_S - I_D - I_{E-EFF} + A \quad (1)$$

in which:

R_O – basic signal-to-noise ratio

I_S – sum of all degradation factors appearing simultaneously with the voice transfer

I_D – represents degradation caused by the voice transfer delay

I_{E-EFF} – factor representing the impairment of quality caused by packet loss

A = advantage factor (from 0 to 20; based on the type of codec)

Previously mentioned recommendation includes also a set of recommended values which enable to simplify the calculation, so that it corresponds with packet networks. Following parameters are important, regarding the QoS in packet networks, namely Network delay, Jitter and Packet loss [6]. These parameters are included in two factors used to calculate the R-factor, in I_D (2) and in above mentioned I_{E-EFF} [4].

$$I_D = I_{DTE} + I_{DD} \quad (2)$$

The parameter I_{DTE} represents the factor of impairment caused by echo (Echo-cancellation has been solved in ITU-T G.168 recommendation), and I_{DD} represents the factor of impairment caused by too long transfer delay. By keeping all the default values during the calculation, the R-factor reaches a final value of 93.35 [1]. In order to meet user's expectations, the value of 70 or more is needed [1]. Simplified calculation of the R-factor uses the following final formula:

$$R = 93.35 - I_D - I_{E-EFF} \quad (3)$$

Except from the R-factor values, a rating scheme of 1 – 5, called *MOS* (Mean Opinion Score), can be used as an evaluation scale. Conversion of the R-factor values to MOS scale values is described in previously mentioned recommendation. Objective evaluation methods for video services are based on mathematical comparison of individual frames and evaluating the similarity between them.

PSNR is defined via the mean squared error (MSE) which represents squared deviation between a tested and an original sample and the maximum possible pixel value of the image as follows: [9]:

$$PSNR = 10 \log_{10} \left(\frac{MAX^2}{MSE} \right) [dB] \quad (4)$$

The SSIM method considers perception of image by human eye. It evaluates the visual impact of brightness shifts in image, variations in contrast and other detected errors of the image in comparison to original. SSIM referential values are on scale 0 – 1; with 0 meaning zero similarity to original and 1 meaning two completely identical images. Final SSIM value is a combination of three parameters, with original signal x and encoded signal y being defined as follows [8, 11]:

$$SSIM(x,y) = [l(x,y)]^\alpha [c(x,y)]^\beta [s(x,y)]^\gamma \quad (5)$$

- Element $l(x,y)$ compares the brightness of signal
- Element $c(x,y)$ compares the contrast of signal
- Element $s(x,y)$ measures the structure of correlation
- $\alpha > 0, \beta > 0, \gamma > 0$ measures the weight of individual elements

Four levels of utilization were used to analyse the dependence of network utilization on the quality of voice and video – 25%, 50%, 75% and 100%. The voice service was represented by two codecs with the highest assessment according to PESQ ranking – G.711 A-law and G.729. The video service was represented by two currently the most used video codecs for digital broadcasting – MPEG-2 and MPEG-4(h.264). Routers were not set to any specific data processing policy (Best Effort). Ethernet bandwidth consumption of voice codec G.711 is approximately 3 times bigger than G.729 (90.4 vs 34.4 kbps). Codec G.729 also contains implementation of PLC algorithm, hence offers better resistance during the higher level of packet loss in network [6]. For the purposes of mathematical formulation of the R-factor values drop with regard to the level of network utilization, it is essential to analyse the ratio of constituent factors I_{E-EFF} a I_D (packet loss and total – mouth to ear – delay) at such occurrence. Following tables gives information about the measurement results.

Table 1. Dependence of network utilization and QoS for the voice

Network utilization	MOS		R-factor	
	G.711	G.729	G.711	G.729
Only 5 calls	4.37	4.03	91.45	81.31
25 %	4.33	4.03	90.22	80.15
50%	4.28	4.03	88.92	80.1
75 %	3.44	3.99	68.49	76.96
100 %	2.92	3.89	56.81	76.08

Table 2. Packet loss during the voice service testing (%)

Codec	Only 5 calls	25 %	50 %	75 %	100
G.711	0.1	0.25	0.41	1.44	2.45
G.729	0.02	0.03	0.15	0.42	0.51

Another group of measurements was carried out to express the I_{E-EFF} factor, using Linux tool *netem*. This tool is capable of setting required packet loss in Ethernet interface of a network adapter. The results of measurements showed following regressive equations:

For codec G.711 PCM

$$Y = \sqrt{(a + (b * X))}, R^2 = 99.37 \% \quad (6)$$

$$Y = I_{E-EFF} \quad X = \text{packet loss } (\%)$$

$$a = -207.44 \quad b = 536.251$$

For codec G.729 CS-ACELP

$$Y = \sqrt{(a + (b * X))}, R^2 = 99.34 \% \quad (7)$$

$$Y = I_{E-EFF} \quad X = \text{packet loss (\%)} \\ a = 91.1429 \quad b = 157.535$$

Following diagrams describe functions of designed models, along with correspondence comparison to results from other papers. According to the diagrams, better results for G.729 can be seen in case of bigger data loss.



Fig. 1. I_{E-EFF} factor as a function for G.711 (left) and G.729 (right)

The second factor I_D , is being expressed by regressive equation used in our previous work and specified at the beginning of the interval, as follows [5]:

For I_D at interval of 0-180 ms

$$Y = (a + (b * X^2))^2 \tag{8}$$

$$Y = I_D \quad X = \text{total delay (ms)} \\ a = -0.606899 \quad b = 0.0000602265$$

For I_D at interval from 180 ms

$$Y = \left(a + \frac{b}{X}\right)^2, \quad R^2 = 99.83 \tag{9}$$

$$Y = I_D \quad X = \text{total delay (ms)} \\ a = 7.87441 \quad b = -1192.81$$

If both of these factors are specified, it is possible to check the equation (3) for realized measures. Table 3 shows that I_D factor consists of several components. Fixed delay (transformation to digital form, creation of voice packet...) reached the average value of 61ms for G.711 and 75 ms for G.729. Variable delay includes the time necessary for packet processing in routers, and variation of Jitter delay [6]. Table 4 describes the results of variable delay for two routers through which the network was utilized, along with I_D a I_{E-EFF} factors. The relative standard deviation (%RSD) is a statistical measure of the precision for a series of repetitive measurements. The RSD is calculated from the standard deviation and is commonly expressed as a percentage (%). Due to the very low calculated values of RSD, it can be pronounced that mentioned regressive equations implemented for the application are truly accurate. An interesting fact is that the delay caused by two routers is quite big (94-38=56 ms) already at network utilization of 50%.

Table 3. List of individual parameters for G.729 from IxChariot programme, acquired during the testing

MOS Average	R-value Average	End-toEnd delay [ms]	One-Way Delay Average [ms]	Jitter Average [ms]
3.82	75.28	218.34	92	51.310

Table 4. Calculation of the R-factor by implementing equations and values obtained by measurements

Network utilization	Codec	Variable delay	End-to-End delay	I_D+I_{E-EFF}	R-factor	%RSD
50 %	G.711	94 (38)	154	4.19	89.16	0.27
	G.729	98 (41)	173	12.14	81.21	1.39
75 %	G.711	120 (49)	181	25.41	67.94	0.8
	G.729	121 (51)	196	15.74	77.61	0.84
100 %	G.711	145 (56)	206	37.49	55.86	1.67
	G.729	146 (55)	221	19.17	74.18	2.5

NOTE: NUMBER IN BRACKETS DETERMINES THE SIZE OF JITTER WITHIN VARIABLE DELAY, DELAY VALUES ARE EXPRESSED IN MILLISECONDS

The conclusion from [6] is confirmed, since it emphasizes the importance of network performance type. During our measurements, we were equipped not only with voice calls in network performance, but also with a streamed video and several UDP flows generated by iperf. Such a huge data extent resulted in decrease of network efficiency already at its half-sized utilization. The delay values on routers acquired by measurements, as well as Jitter value, will be applied for the calculation of total delay with regard to number of used routers and network utilization at individual sections expressed by percentage (0% 50% 75% 100%). For calculation of buffer size to eliminate Jitter delay, minimally 1.5 times larger size of buffer, compared with measured average delay, is required [12]; the application is based on this recommendation. In order to use an adequate delay value on routers with regard to network section utilization, a simple equation with measured values can be implemented for the application:

$$R_D = 44.5963 + 45.3578*(X)^2 \quad (10)$$

In which R_D stands for delay on the two routers and X stands for utilization of the section (0.5; 0.75; 1). When implementing QoS policy other than Best Effort, results of measurements are processed [5]. The last delay component which needs to be focused on is Jitter, or more precisely the buffer size on receiving side. In case the buffer is set on low level, its fast filling is usually followed by packet drop. On the other hand, setting the buffer on too high level causes a disproportionate increase of one-way delay. The application based on previously mentioned recommendation and results acquired by measurements calculates the appropriate buffer size. It also offers an option to set any buffer size, however, it is necessary to consider this factor when making an overall evaluation of final quality of service. For the purposes of such an evaluation, following Table 5 was compiled, using the results from [7].

Table 5. Impact of buffer size

Jitter [ms]	Buffer size [ms]	Average buffer loss [%]
20	40	0.1
20	60	0.03
40	40	5.98
40	60	1.45
80	40	14.75
80	60	7.9

The Table 5 shows that if the buffer is more than double-sized, there is a minimal chance of packet loss increase. On the contrary, if the buffer size is approximately the same or smaller than average Jitter, it results in significant packet loss increase. The ratio between Jitter and buffer size may be expressed as follows:

$$P_{lb} = -4.65399 + 9.71812 * X \quad (11)$$

The result P_{lb} represents the increase of packet loss in network (%) and X represents the ratio between Jitter and buffer on the scale 0.5 – 2. The results of quality of video measurements stress an enormous impact of packet loss in network. Since streamed broadcasting via the Internet known as IPTV is a one-way service, the delay does not play such a significant role in this case. Newer codec MPEG-4 uses stronger compression than its predecessor.

Table 6. Dependence of network utilization and QoS for video service

Network utilization	MPEG-2		MPEG-4 (h.264)	
	PSNR [dB]	SSIM	PSNR [dB]	SSIM
Only 1 stream	51.835	0.989	51.235	0.989
25%	49.124	0.985	46.511	0.965
50%	47.16	0.976	32.001	0.949
75%	40.72	0.961	24.531	0.919
100%	27.234	0.929	17.036	0.763

While streaming the video that demands much higher capacity than voice, more significant packet loss was detected. The values of packet loss for both, MPEG-2 and MPEG-4, expressed in percentage, can be seen in Table 7.

Table 7. Packet loss values during the testing of video service (%)

Codec	Only 1 stream	25 %	50 %	75 %	100
MPEG-2	0	0.06	0.1	0.31	4.78
MPEG-4	0	0.05	0.16	0.36	9.45

The decrease in QoS, altogether with the analyzed impact of packet loss on final quality of video, was processed into following regressive equations in which X stands for percentage packet loss ratio in network. Verification of these mentioned formulas are published in our paper [5].

Table 8. Regressive equations of video service [5]

	MPEG-2	MPEG-4 (h.264)
PSNR	$Y = \frac{1}{a + b\sqrt{\ln(X)}}$	$Y = a + b * \ln(X)$
SSIM	$Y = \frac{1}{a + b\sqrt{\ln(X)}}$	$Y = \sqrt{a + b * \ln(X)}$

Table 9. Coefficients of regressive equation models for video service [5]

Y	Codec	Coefficient		R ²
		a	b	
PSNR	MPEG-2	0.0297642	0.00444535	97.37
PSNR	MPEG-4	22.19	-2.29442	92.85
SSIM	MPEG-2	1.03622	0.011083	98.01
SSIM	MPEG-4	0.81727	-0.10898	92.35

3.2 Features of Implemented SW Tool Designed According to Simulation Model

Since the application focuses on the whole scale of Triple-play services, there is one service left – data. We have dealt with this service as well in previously mentioned paper [5], which detected the ratio between network capacity and actually reached speed of ftp service with regard to utilization and implemented QoS policy. Overall application settings:

- Calculation of MOS and the R-factor, based on prediction of total delay, Jitter buffer, selected packet loss and network utilization
- Calculation of SSIM and PSNR for video with regard to selected packet loss in network
- Bandwidth for data service within Triple-play

Following parameters are possible to set for the purposes of application settings specification:

- Bandwidth uplink and downlink at end user
- Number of routers, their configured policies and utilization of individual sections
- Distance between communicating points

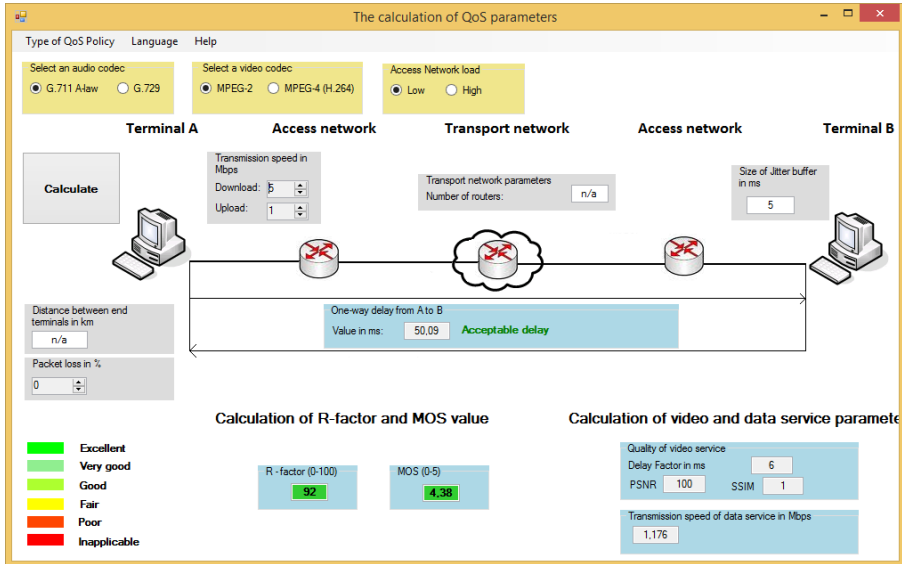


Fig. 2. Graphical design of the application

The application offers as next optional parameters the total delay, De-Jitter buffer size and packet loss values (if not filled then default values are used: 0, 60, 0).

4 Conclusion

The aim of this paper was to bring new simulating model able to predict objective QoS parameters of Triple play services in IP network. The accuracy of mathematical models was verified through a method of comparing the calculation with results of actual experiments. Measured results then served as predefined scenarios concerning packet loss in network, delay at specific policy implementation and selected network utilization. The use of proposed models provides the application with an ability to react immediately to any changes of default settings and adapt to a specific status as much as possible. Further improvement of the application is expected.

The next step should be an analysis of the impact of security and coding mechanisms implemented on QoS parameters. Security is a highly discussed topic nowadays, and protocols such as IPsec, SSL/TLS, SRTP or ZRTP are becoming more and more frequently used to secure the content of voice or video. Therefore, the proposed analytical computational model should take into account various security measures in network as well.

Acknowledgement. This work was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070) and by the Development of human resources in research and development of latest soft computing methods and their application in practice project (CZ.1.07/2.3.00/20.0072)

funded by Operational Programme Education for Competitiveness and partially was supported by Grant of SGS No. SP2014/72.

References

1. Voznak, M.: E-model modification for case of cascade codecs arrangement. *International Journal of Mathematical Models and Methods in Applied Sciences* 5(8), 1439–1447 (2011)
2. Cole, R.G., Rosenbluth, J.H.: Voice over IP performance monitoring. In: *ACMSIGCOMM Computer Communication*, New York (2001)
3. Managing Voice Quality with Cisco Voice Manager (CVM) and Telemate, http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_tech_note09186a00800946f8.shtml
4. ITU-T G.107, The E-model, a computational model for use in transmission planning, ITU-T Recommendation G.107, ITU-T Geneva (2010)
5. Frnda, J., Voznak, M., Rozhon, J., Mehic, M.: Prediction Model of QoS for Triple Play Services. In: *21st Telecommunications Forum TELFOR 2013* (2013)
6. Karam, M., Tobagi, F.: Analysis of delay and delay jitter of voice traffic in the Internet. *Computer Networks* 40, 711–726 (2002)
7. Kovac, A., Halas, M.: E-model mos estimate precision improvement and modelling of jitter effects. *Advances in Electrical and Electronic Engineering* 10(4), 276–281 (2012)
8. Changhoon, Y., Bovik, A.C.: Evaluation of temporal variation of video quality in packet loss networks. In: *Signal Processing: Image Communication*, vol. 2011, pp. 34–38 (2011)
9. Feamster, N., Balakrishnan, H.: Packet Loss Recovery for Streaming Video. In: *12th International Packet Video Workshop*, Pittsburgh, PA (2002)
10. Molnar, K., Vrba, V.: DiffServ-based user-manageable quality of service control system. In: *7th International Conference on Networking, ICN 2008, Cancun, Mexico*, Article number 4498208, pp. 485–490 (2008)
11. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multi-scale structural similarity for image quality assessment. In: *Proc. 37th IEEE Asilomar Conference on Signals, Systems and Computers* (2002)
12. ITU-T G.113, Transmission impairments due to speech processing, ITU-T Recommendation G.113, ITU-T Geneva, Switzerland (2001)

Study on Security Analysis of RFID

Yi Hou¹ and Jialin Ma²

¹ Shenyang Radio and TV University, 110003, Shenyang, China

² Shenyang Normal University, 110034, Shenyang, China

jialinma@126.com

Abstract. This document is in the required format. The specialties and limitation of RFID system and device will incur lots of security problems especially when the system contains a lot of tags and many readers, the traditional methods become helpless. Here we present and analyze an improved hash key technique based on multiple readers.

Keywords: RFID Security, Reader, tag, Hash Algorithm.

1 Introduction

RFID (Radio Frequency Identification) is to become a reality, because modern supply chains require higher performance, greater information flow and flexibility. RFID products have been widely used in the retail, pharmaceuticals, transportation, defense, and with packaging consumer goods industry. RFID technology can improve the forecast level and container security, Reduce the occurrence of the problem of cargo shrinkage, recall and stock at the same time, thus saving billions of dollars in the cost of the supply chain. Wal-Mart, Albertson's and Best Buy and other retailers have asked their top suppliers to all enabled RFID in late 2005. These requirements extended to thousands of suppliers, stores and distribution centers, and they need to assemble a multifunction distributed reader, which raises several technical challenges. Now, the majority of enterprises using RFID technology are concentrated in a small-scale pilot and one-stop concept validation. However, in order to realize the benefits of RFID, the technology must be adopted in the enterprise-wide.

2 RFID Security Threats

RFID collection of information about the personal data, will likely be used by hackers, retailers, and government use of which will result on RFID security and privacy threats.

2.1 Eavesdropping

Tag is designed to be as long as the RFID reader can be found the read mode, in general, whether the label can be read, the main control over the RFID reader itself,

when RFID reader want to read the tag data, labels are not refuse to read RFID reader. Therefore, the label is read, the holder of the label does not know at this time the tag is being read or has been read, and this kind of read in a long distance only likely to occur. If we use RFID in the passport, the cause of the problem is a person know the identity of this person through an RFID reader, and this issue has serious implications for individual privacy and security.

2.2 Tracking

RFID reader can record the tag's unique identification code, so if it is used on the battlefield, it will be revealed label holder, and risk the users. The other hand, if the application in the usual life such as cellular phones, then the user of these products will have the possibility of being tracked , as long as someone was through the RFID reader fully meeting the above mentioned behavior patterns.

2.3 Fake

Forgery here means an attacker can correctly write the RFID reader reads the text of the label format, and further deceive the RFID reader, so as to achieve the purpose of forgery. For example, with stored value cards of shopping malls, someone write a fake label on the label through this way, to set the amount in this tag to 30 million yuan, while the original label value of 100,000 yuan and then use the label (fake Tag) instead of the original label to achieve the unlawful purpose of deception.

2.4 Replay Attacks

Since the transmission between the RFID reader and Tag is the use of radio waves transmitted, so Challenge and Response may cause data between the two easily intercepted. Because message is easy to be intercepted when transmitted, the cause the interception can modify the content of the captured, which will produce a great danger for all applications of RFID products industry. We can use re-encryption of transfer data, to make the one given the data can not know the data content.

2.5 Denial of Service

The main intention is to say, blocking requirements between the RFID reader and Tag the reply (Challenge and Response). Because the label can not resist the RFID reader to read the label, when the label received, this request will return the data to the tag RFID Reader. But to consider is users held label may not want to reply to the request of the RFID Reader, because sometimes such a request is not normal requirements, while illegal RFID Reader want to read your personal data in label, so this preventive action is very important. Japan has developed a similar card case box label in the box, RFID Reader radio waves requests from labels can be blocked. Therefore, as long as

the label is in the box, it will not be any RFID Reader to read into the data. When to use RFID cards, and then label from the box can be out.

To resolve the security issues, a variety of programs, including to kill killing label tag Faraday mesh cover of Faraday net cover, active jamming, smart tags, block tags and Hash lock.

3 Solutions of RFID Security Issues with Traditional Hash Lock

3.1 Hash Lock

Hash lock is a enhancement technology of boycott of the unauthorized label access privacy. The label verify the reader works as follows, each key K of tag is stored in readers, corresponding tag storage metaID, wherein metaID the = Hash (K). Tag receives the request of reader to read and sends metaID as response, and then reader queries the tag and obtains label metaID corresponding key K and sends the label. Label sends a key K with Hash function reader to check Hash (K) is the same with metaID or not. If same, it will unlock the real ID of the transmitting tag to the reader.

The program presented a low-cost solution to the security and privacy issues. We only need a Hash equation and store metaID value. But it can not prevent to be tracked because of its fixed metaID as well as RF tag's response can be predicted in advance, and accessing key k and labeling ID all through the front channel, so it is easy and can be the enemy eavesdropping.

3.2 Random Hash Lock

In order to avoid being tracked, the reaction of the radio-frequency tag can not be predicted but random. As an extension of the Hash Lock, Random Hash Lock solves the label positioning Privacy. Random Hash Lock scheme makes the output information of label readers visit each time different.

Random Hash Lock works as follows. The tag contains Hash functions and random number generator, and the back-end database to store all label ID. After reader requests accessing to labels and label receives access request, Hash function calculates tag ID and a random number r (generated by a random number generator) and r 's Hash value (Hash ID i || r). Label sends (r , Hash (ID i || r)) data to the reader in request, and reader sends to the back-end database, backend database exhaustive search for tag ID and all Hash value of r . If Hash (the ID j || r) = Hash (ID i || r), ID j is the ID of the corresponding label. The label receives ID j the reader sent and unlocks.

Hash functions can be done in the case of low-cost, but you want to integrate the random number generator to limited passive tags of low-cost computing power, is very difficult. Secondly, Random Hash Lock solwew only label positioning privacy issues. Once labeled secret information is intercepted, the enemy can obtained the access control rights information through backtrack label history to infer tag holder privacy. The program also exists the security issues of long distance to read the

channel transmission and ID can be easily intercepted. Decoding operation of the back-end database is through an exhaustive search, and all tags need exhaustive searched and Hash function calculated.

4 Hash Lock Method Based on the Multi-reader's Random Read Control

The Hash Lock increased atresia and unlock status with simple Hash function and access to the communication between the tag and reader. But it can not solve the location privacy problem and man-in-the-middle attack. Random read control of Hash Lock method to avoid the disadvantage of being tracked, however, this method is only suitable for the user of a small amount of radio-frequency tag. When the system contains a large number of tags and readers, the Hash Lock and Random Hash Lock method are on the powerless. This paper presents an improved method based on multi-reader Hash lock to solve this problem.

4.1 The Structure of Every Part of the RFID System

RF label consists of two parts: one part is the read-only memory (ROM) and random read memory (RAM). ROM stored the Hash value of the tag ID, and the RAM storage is the ID of identified readers. Another part of the logic circuit, mainly for some simple calculations, such as to calculate Hash equations or simple pseudo-random number.

Reader communicates with radio frequency tags in wireless. When the system has more than one reader, every reader has ReaderID to identify the number of identified reader. For example, all readers in a supermarket have the same ReaderID, suggesting that they are derived from the supermarket. When the reader issues access request to radio frequency tags, the tag test reader through the reader ReaderID. Readers connect and communicate with a back-end database at the same time identification tag and run the application.

The back-end database stores a pair of RF tag ID and Hash equation values: [TagID, hash (TagID)]. In general, the back-end database and readers connect through wired and safe passage.

4.2 Working Principle

Reader needs to query the RF tag ID, and when the system has more than one reader, they may have same or different ReaderID. When the readers have different ReaderIDs, the labels need to first determine whether the reader is authenticated. If the reader is certified, the label is in response to the reader and the reader to obtain ID. Before readers are responded and received tag ID information, the reader and the label to determine the certification system. Because the reader ReaderID is stored in the tag RAM in advance, the label identifies permission reader with the reader ID. The label does not respond readers with no permission. Therefore, it is impossible to be tracked

by enemy (because the reader has permission). In addition, this kind of the privileges process is based on random number generated by the tag, which can also prevent the enemy coaxing.

When the radio-frequency tag receives the request of the reader, the radio-frequency tag first generates a random number k and sends the reader. The reader receives and sends it to the backend database, and back-end database calculates a $(k) = \text{Hash}(\text{ReaderID} \parallel k)$. Then a (k) is sent to readers, and the reader transmits a (k) to the label. RF tags computes the Hash $(\text{ReaderID} \parallel k)$, and the tag comparison reader calculates whether the Hash is equal with the label a (k) values. If equal, the reader passes the certification and label sends some TagID related information to it. If not, the reader is not certified to be shielded.

After the certification of the reader, the RF tag reader response authentication Hash (TagID). When the reader receives data the Hash (TagID) value, it will communicate with the back-end database and find (TagID, Hash (TagID)). The reader will get corresponding TagID. Even if Hash (TagID) value is eavesdropping, when the label sends out its value, the eavesdropper will not know the value of Hash. Because the eavesdropper can not determine the relationship between the TagID with Hash (TagID).

When RF tag memory updates certification reader ReaderID, and an object is transported from one warehouse to another one, the reader certified will change from previous warehouse reader to this reader. Reader Hash (Tag ID) value passed to the back-end database, the database notice ReaderID stored in the tag to update. Accordingly, the database find out the New ReaderID and pass it on to the reader. When the reader receives New ReaderID, the reader make the value "XOR" with Old ReaderID, and send the value after 'XOR' to the RF tag.

The tags can derive New ReaderID from the exclusive-OR value and Old ReaderID, and at last ReaderID can be updated. In ReaderID update process, even if the XOR value is leaked, the enemy can not get New ReaderID. Because they can not get Old ReaderID, we prevent the coaxing.

4.3 Method Analysis

In the certification process, even if the enemy eavesdrop the reader output a (k) , they can not obtain certification in the next step. Because a (k) value in each of the authentication process needed is changed. The former certified a (k) value is meaningless for the latter certification. After Authentication, the RF tag output the Hash (TagID) instead of TagID. Hash equation is difficult to find the inverse function, so enemy can not obtained TagID value even have captured output Hash (TagID) value. When the RF label need to update ReaderID of memory, the updated ReaderID encrypt after old ReaderID to prevent eavesdropping.

In short, even if the above method in the communication between the reader and the RF tag encountered enemy eavesdropping, it's also safe.

RF tag is shield to enemy, and only response to certified reader. Moreover, as mentioned above, the enemy can not forge certification reader. Because there is no

label output, the enemy can not track tags to keep track of what customers pay for or buy. Location privacy and items guests bring have been protected.

The improved method has high operation speed, and low cost. When identifying one from N radio-frequency tag known, the reader need only perform the Hash operation one time and ID search N times. Random read control Hash lock method requires at least N times of the Hash operation and N ID search. Obviously, at the same security level, the improved method proposed for computing load significantly reduced. And the authentication process is dependent on the N -known tag ID and a hash lock equation, therefore, with the increase in the number of tags, the computing load increases slowly.

Because the calculation load is low, and slowly increased with the increase in the number of the radio-frequency tag, the method is very suitable for a RFID system with a large number of labels and protected.

5 Conclusion

A random Hash lock improved method proposed above is suitable for system containing multiple reader and tag, such as logistics management. The method has the advantages of high security, low load, applied to the occasion containing a large number of label application, to solve the location privacy and man-in-the-middle attack problems, even if the enemy has stolen label output, but they can not get the tag ID. However, in this scenario, the execution of one time certification requires for twice Hash calculation of Tag, which corresponding increase some cost, and this is also its disadvantages.

Because of the growing attention to RFID, human life and behavior will be easier to be tracked through the arrangement of a large number of RFID readers. I believe that security will be one of the RFID focus of future development and the success factor.

References

1. Zhou, Y.-B., Feng, D.-G.: Design and Analysis of Cryptographic Protocols for RFID. *J. Chinese Journal of Computers*, 23–25 (2006)
2. Riebackmr, Crispo, B., Tanenbaum, A.S.: The evolution of RFID security. *IEEE Pervasive Computing*, 62–69 (2006)
3. Su, W., Cui, Z., Wang, X.-J.: Research on Hash chain-based RFID privacy enhancement tag. *J. Computer Applications* (10) (2006)
4. Thornton, F.: *Syngress RFID Security*. Syngress Publishing, Canada (2006)
5. Sarma, S.E., Weis, S.A., Engels, D.W.: Radio-frequency identification: Secure risks and challenges. *RSA Laboratories Cryptobytes* (2003)
6. Castelluccia, C., Avoine, G.: Noisy Tags: A Pretty Good Key Exchange Protocol for RFID Tags. In: Domingo-Ferrer, J., Posegga, J., Schreckling, D. (eds.) *CARDIS 2006*. LNCS, vol. 3928, pp. 289–299. Springer, Heidelberg (2006)

Web Services Discovery with Semantic Based on P2P

Jin Li, Yongyi Zhao, and Bo Song

College of Software Shenyang Normal University Shenyang, Liaoning Province, China
lijin4407@sina.com

Abstract. Web services provide a loosely coupled paradigm for distributed processing. To reduce the complexity and improve the compatibility with the legacy system, Web services usually provide a centralized service discovery and management mechanism. But there is a performance bottleneck and single-point of failure in this centralized method. To achieve the high scalability and efficiency, the decentralized Web services discovery approach based on P2P can be used. This paper presents a new model for structured Web services organization based on P2P. First, establish the semantic tree in which semantic information is contained in Web services addressing, routing table, and routing algorithm. Semantic tree achieve the semantic aggregation and routing process of Web services. Second, implant the Web services node in the semantic tree and the services computation domain is formed. Based on the structured P2P, Web services can be scheduled in varied collaborative computing model, which decentered and improve the performance.

Keywords: Web services, peer-to-peer, Web services discovery, semantic Web.

1 Introduction

Web service is defined as service oriented architecture [1]. In Web services architecture [2], all functions are defined as independent services that can be invoked with a well-defined interface. We can call these services to perform the business process. Web services are self-describing software applications that can be advertised, located, and used across the Internet using a set of standards such as SOAP, WSDL, and UDDI [3]. Since it is text-based and self-describing, SOAP messages can convey information between services in heterogeneous computing environments without worrying about conversion problems, there are many other Web Service specifications. Two of them, which are based on XML, are Web Service Description Language (WSDL) and Universal Description, Discovery and Integration (UDDI) [4]. WSDL defines a standard method of describing a Web Service and its capability, and UDDI defines XML-based-rules for publishing Web Service information. Messages are exchanged through the SOAP protocol. This allows the data to be exchanged regardless of where the client is in the network.

The shortcoming of the current UDDI model is that it limits the service discovery to functional requirements only. It is true that there may be more than one Web

services available that can meet the functional requirements with different quality of service attributes. Therefore the ability of incorporating quality of service into service discovery process becomes very important.

The increasing number of web services demands for an effective solution to look up and select the most appropriate services for the requirements of the user. Web services discovery is the process of finding an appropriate service provider for a service requester through a middle agent [5]. First, service providers advertise their capabilities to middle agents, and middle agents store this information. Second, a service requester asks a middle agent whether it knows of service providers' best matching requested capabilities. Finally, the middle agent tries to match the request against the stored advertisements and returns a subset of stored service providers' advertisements. But the traditional web service discovery mechanism is based on the centralized UDDI, which lead to performance bottlenecks if a large number of clients visit it. And there is also a potential single-point fault because of the centralized UDDI[6]. The registry once failure, the whole service discovery will not be able to carry on. Using P2P completely distributed advantages, we can solve the problem [7-10].

Peer-to-peer (P2P) refers to a class of systems and applications that employ distributed resources to perform a function in a decentralized manner. The advantages of P2P technology embodied in the respects: decentralized, scalability, robustness, and high performance [11]. With the use of P2P mechanism as the service repository network, the Web services discovery system is highly scalable in terms of number of registries and services. There are four kinds of topological form in P2P system according to the structural relationship between the nodes.

In centralized topology, the central server is easy to cause the collapse of the entire network, so the reliability and security is low. The unstructured network cover is a completely random graph, the link between nodes are not predefined topology. The network bandwidth consumption is very large in such Flooding query system and the performance generally cannot be guaranteed. DHT network mainly adopts distributed hash table technology to organize the nodes. The adoption of certainty topological structure, DHT can provide accurate findings. Hybrid Structure absorbs the advantages of centralized structure and fully distributed unstructured topology. Select the high performance nodes as super nodes and the information of other nodes is stored in super nodes. Query request is only forwarded between the super nodes and super nodes forward the request to the appropriate leaf nodes. As the result of the super node index function, the search efficiency is greatly increased. So the Hybrid Structure has the advantage of performance, scalability and easier to manage, but it has also been affected in fault tolerance and it is vulnerable to be attacked because of the dependence of super Node. Because the DHT network has the good performance in scalability, reliability, maintainability and query efficiency, it is adopted in Web services system.

2 System Design and Implementation

2.1 Web Services Description with Semantic

Web service is a function module which can be called each other across the network. It can accept the request that come from the remote client or other Web service and return the processing result. Web services can be defined by six attributes: name, address, interface, implement, semantics, and constraints.

Web service=<name, address, interface, implement, semantics, constraints>

Semantic is the valuable characteristic about Web service itself or the host environment. Web services must be located that might contain the desired functionality, operational metrics, and interfaces needed to carry out the realization of a given task. The semantic description of Web services allows more accurate searches and supplying a better solution for the selection, composition and interoperation of Web services. According to the associated between Web services and semantic, semantic can be classified as group semantic and individual semantic. We can use the group semantic classifying and forming the Web services collections. Accordingly, the semantic that do not belong to the group semantic is individual semantic. There may be a containment relationship between different group semantics. We can extract the group semantic between Web services and establish the semantic tree to describe the Web services architecture. Semantic tree is an undirected tree, in which each node shows the group semantic of Web services. Web services can be categorized on the basis of the group semantics, the father-son relationship is equivalent to the inclusion relation between group semantics. Figure 1 shows a semantic tree of Web services on the basis of the group semantics.

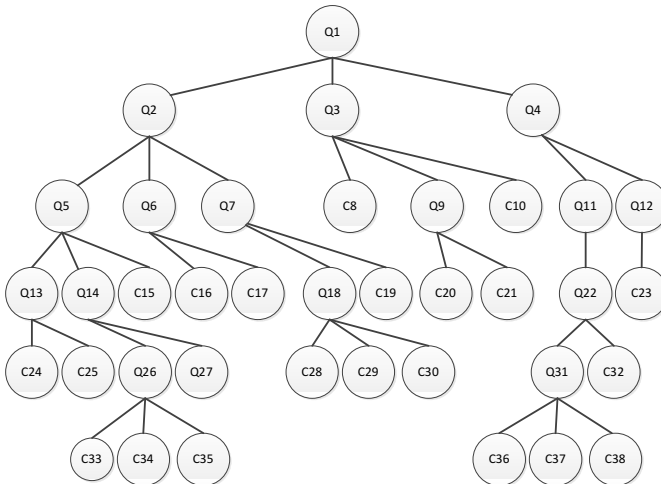


Fig. 1. A semantic tree of Web services

Web services are in the leaf nodes and other nodes are the query nodes in the above semantic tree[12,13]. In DHT network, each node must be assigned identification when the nodes join the network. The identification determines the routing between nodes by with one node can find another node. The node addressing is the Web service addressing in semantic tree. The traditional DHT network, the digital or binary code generated randomly or sequentially increased can be used as service identification. The method lacks of considerations about Web service semantic, so the Web service node is no rule in distribution. In order to solve the above problem, the Web service can be describe using the semantic vector. Some semantic characteristics of Web services such as function, performance and management can be added in the process of addressing. Web services id can be defined:

$$Id = \langle fword1, fword2, \dots, fwordn, ncode \rangle$$

fword1, fword2, ... , fwordn is the semantic sequence and ncode is the name sequence.

The coordinate of nodes in the semantic tree can be described using the triples.

$$Coordinate = (L, R, N)$$

L describes the serial number of layers in the semantic tree and R is the serial number of feature regions. The nodes that have the same parent node have the same feature region. N is the inner number in a feature region. The coordinates of the nodes in the semantic tree are shown in figure 2.

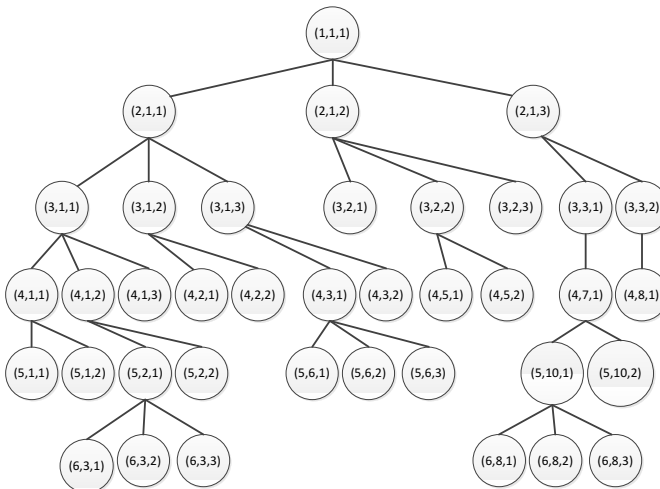


Fig. 2. The coordinates of the nodes in the semantic tree

2.2 The Routing of Web Services with Semantic

In accordance with the information stored and searched in the node, Web services discovery system based on P2P can be divided into two categories: unstructured and structured system [14]. Each node storages specific information or index in structured

P2P system. The neighbor nodes have been well defined in the structured P2P networks, so it can avoid the flooding search that be used in the unstructured P2P system. And measurement studied show that unstructured P2P system does not scale well because of the large volume of query messages generated by flooding [15]. By contrast, structured P2P networks such as those using distributed hash tables (DHT) [16] maintain a structured overlay network among peers and use message routing instead of flooding. DHT technology can solve the Web services positioning and search in P2P network through the DHT layer that between the network application layer and the network routing layer. Each Web service maintains a routing table in which the information of other Web services is stored. Web services are able to visit each other using the information and the overlay network can be formed finally. Item is the routing information in a routing table by which other Web services can be routed. The items in the routing table are divided into multiple routing buckets. The routing buckets are numbered by sequence, expressed as bucket1, bucket2, bucket3 and so on. The first component in Web services identifications is the feature value which describes the semantic space of the root. All the Web services identifications must have the same first component in the Web services system. Web services information that the second component is different are stored in bucket1. Similarly, Web services information which is in I+1 layer semantic space are stored in bucket that numbered I. The division of routing buckets and structure of the semantic tree are show in figure 3.

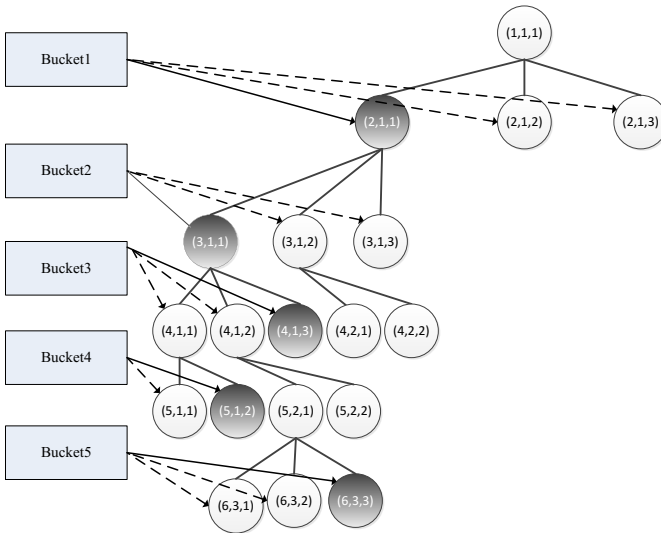


Fig. 3. The division of routing buckets and structure of the semantic tree

2.3 The Design and Implementation of Web Services with Semantic Discovery Algorithm Based on P2P

1. Routing algorithm

Routing algorithm receives the routing request and returns the most matched next-hop nodes. At the same time extract information from the request and update the routing table by calling routing update algorithm. Firstly algorithm calculates the semantic distance between the request service and the current service. If the semantic distance is 0, the current service is the request service. Otherwise, the routing algorithm returns the highest satisfied services in the routing bucket that matching the request.

```
BEGIN
  IF distance (current node, request service) =0 THEN
    process the request ;
  END
  bucket
  FOR each item DO
    calculate the satisfaction of routing ;
  END
  hops;
  routing update algorithm;
  RETURN hops;
END
```

2. Web services discovery algorithm

After receiving the query request, Web services nodes submit the routing request from themselves as the starting point. Generate and submit the next routing request using the next-hop nodes and loop the process. If the queried service is found, the routing list is return. Otherwise the query is failure.

```
BEGIN
  request
  h_hops
  c_hops
  nx
  c_hops←nx
  WHILE c_hops not null , DO
    n_hops←null
    FOR nx← Web services in c_hops , DO
      hops←the nex-hop list
      n_hops←hops
      h_hops←nx
    END
    FOR each service in n_hops , DO
      distance
```

```

        IF distance=0 THEN
            RETURN nx
        END
    c_hops←services that is not in h_hops
END
RETURN null;
END

```

3 Conclusion

Web services provide a loosely coupled paradigm for distributed processing. Web services discovery is the most important component in the Web services architecture, which has turn to an arduous task. But the traditional Web services discovery mechanism is based on the centralized UDDI, which lead to performance bottlenecks if a large number of clients visit it. Because of the peers self-organization, P2P system implements the scalability, fault resilience, intermittent connection of Web services. Web services discovery base on P2P tends to balance the load on the system, robust and efficient. Each Web service is one node in the overlay network based on P2P. Web services nodes are organized as a semantic tree. By embedding the semantim information into the Web services addressing, routing table and routing algorithm, Web services implement the aggregated distribution and routing process by the semantic. This kind of organization model and routing algorithm have the fast convergence speed and some bandwidth are saved.

Acknowledgment. This paper is supported by the institution of educational science Research plan of Liaoning, No. JG11DB247.

References

1. Lublinsky, B.: Defining SOA as an architectural style (EB/OL), <http://www.ibm.com/developerworks/webservices>
2. Haas, H., Brown, A.: Web Services Glossary (EB/OL) (2004), <http://www.w3.org/TR/ws-gloss/>
3. Papazoglou, M.P., Georgakopoulos, D.: Serive-Oriented Computing. *Communications of the ACM* 46(10), 25–65 (2003)
4. UDDI (EB/OL) (2013), <http://uddi.org/>
5. Lublinsky, B.: Defining SOA as an architectural style (EB/OL), <http://www.ibm.com/developerworks/webservices>
6. Dooley, K.: *Designing Large Scale Lans*. O'Reilly Media (2009)
7. Harrison, A., Taylor, I.J.: WSPeer-An interface to web service hosting and invocation. In: *Proceedings of 19th IEEE International Parallel and distributed Processing Symposium, IPDPS 2005*, p. 1420050 (2005)

8. Sahin, O.D., Gerede, C.E., Agrawal, D.P., El Abbadi, A., Ibarra, O.H., Su, J.: SpiDeR: P2P-based web service discovery. In: Benatallah, B., Casati, F., Traverso, P. (eds.) ICSOC 2005. LNCS, vol. 3826, pp. 157–169. Springer, Heidelberg (2005)
9. Verma, K., Sivashanmugam, K., Sheth, A., et al.: METEOR-S WSDI: A Scalable P2P Infrastructure of Registries for Semantic Publication and Discovery of Web Services. *Information Technology and Management* 6(1), 17–39 (2005)
10. Bianchini, D., De Antonellis, V., Melchiori, M., Salvi, D.: A Semantic Overlay for Service Discovery across Web Information System. In: Bailey, J., Maier, D., Schewe, K.-D., Thalheim, B., Wang, X.S. (eds.) WISE 2008. LNCS, vol. 5175, pp. 292–306. Springer, Heidelberg (2008)
11. Milojicic, D.S., Kalogeraki, V., Lukose, R., et al.: *Peer-to-Peer Computing*. HP Laboratories, Palo Alto (2002)
12. Technical Committee on Service Computing (EB/OL) (2013), <http://tab.computer.org/tcsc/>
13. SCC 2004 (EB/OL) (2004), <http://conferences.computer.org/scc/2004>
14. Crowcroft, J., Pias, M., et al.: A Survey and Comparison of Peer-to-Peer Overlay Network Schemes. In: *Submission to IEEE Communications Tutorials and Surveys* (2004)
15. Saroiu, S., Gummadi, P.K., Gribble, S.D.: A Measurement Study of Peer-to-Peer File Sharing Systems. Presented at *Multimedia Computing and Networking* (2002)
16. Ratnasamy, S., Stoica, I., Shenker, S.: Routing algorithms for DHTs: Some open questions. In: Druschel, P., Kaashoek, M.F., Rowstron, A. (eds.) IPTPS 2002. LNCS, vol. 2429, p. 45. Springer, Heidelberg (2002)

Analysis and Enhancement of TCP Performance in Ad Hoc Wireless Networks

Li Miaoyan* and Zhou Chuansheng

Software College, Shenyang Normal University, 110034 Shenyang, China
{lillian1979, jasoncs}@126.com

Abstract. Mobility in ad hoc networks causes frequent link failures, which in turn causes packet losses. TCP attributes these packet losses to congestion. This incorrect inference results in frequent TCP re-transmission time-outs and therefore a degradation in TCP performance even at light loads. In this work, a new TCP enhanced scheme for mobile ad hoc networks is presented. The key design novelty is to identify the network states based on end-to-end measurements in TCP receiver, which successively dispatches the feedback on current network states to TCP sender via the ACK packets. Consequently, the sender takes corresponding measures. Moreover, this paper investigates the reasons of TCP instability in static ad hoc wireless networks and proposes schemes combining MAC layer and route layer enhancements to solve this problem.

Keywords: Ad Hoc Networks, TCP Performance, End-to-End measurements, Instability.

1 Introduction

A mobile ad hoc network (MANET) [1] is a self-organizing system of mobile routers connected by wireless links. In such networks, mobile users may exchange data messages and access the Internet at large. TCP has been the predominant transport protocol used in the wired Internet to deliver data; consequently, numerous Internet applications have been developed to run over TCP. However, TCP performs poorly in ad hoc wireless networks as demonstrated in Ref. 1. The main reason for this poor performance is a high level of packet losses and a resulting high number of TCP re-transmission time-outs.

Various approaches have been proposed to improve TCP performance at the transport layer [2-4]. In Ref. 2, explicit link failure notifications are used to freeze TCP state upon the occurrence of a route failure. Explicit route establishment notifications are used to resume TCP transmissions. A fixed-RTO approach is proposed in Ref. 3 to deal with packet losses due to link failures and route changes. In Ref. 4, a new transport layer protocol that is based on end-to-end rate control is proposed.

* Corresponding author.

Various mechanisms have also been proposed to improve TCP performance at the routing layer [5]. These are sound techniques to improve TCP performance in ad hoc networks, but they rely on global deployment at every node. These techniques may also be difficult to adopt in practice because of the potential heterogeneity of participating nodes.

In contrast, the end-to-end approach is easy to implement and deploy, requires no network support, and provides the flexibility for backward compatibility. In this paper, we explore an end-to-end approach to improve TCP performance in mobile ad hoc networks. We implement our design only at the two end hosts, and do not rely on any explicit network notification mechanism. End-to-end measurements are used to detect congestion, route change, and channel error, and each detection result triggers corresponding control actions.

TCP instability in ad hoc wireless network is also a well-known problem. This paper investigates the reasons of TCP instability in static ad hoc wireless networks and proposes schemes combing MAC layer and route layer enhancements to solve this problem.

2 Related Work

This section describes the work mechanism of ADTCP, and discusses its advantages and disadvantages, because the TCP enhanced scheme proposed in this paper is based on ADTCP.

ADTCP [5] is a typical example of end-to-end protocols. The key design novelty of ADTCP is to perform multi-metric joint identification for packet and connection behaviors. It can detect and classify four network states: Congestion (CON), Channel Err (CHERR), Route Change (RTCHG) and Disconnection (DISC).

ADTCP uses four metrics that tend to be influenced by different conditions to identify the network states: they are Inter-packet delay difference (IDD), Short-term throughput (STT), Packet out-of-order delivery ratio (POR) and Packet loss ratio (PLR). Both IDD and STT are used to estimate CON; POR and PLR estimate RTCHG and CHERR respectively. The mapping between the network states and the metrics is shown in Table 1 (“High” means its value is within the range of the former 30% in descending order; “Low” means its value is within the range of the later 30% in descending order; *: do not care).

Table 1. Metrics patterns in 5 network states

	IDD and STT	POR	PLR
CON	(High, Low)	*	*
RTCHG	NOT(High, Low)	High	*
CHERR	NOT(High, Low)	*	High
DISC	(* , ≈ 0)	*	*
NORMAL	Default		

The necessary condition of congestion state is that both IDD and RTT are CON, and the sign of CON is that the value is HIGH. Given a set of history records we propose a simple density-based technique (RSD-Relative Sample Density) to infer whether a sample value is HIGH or LOW. By dynamically maintaining the sample space, the current metric is compared with the value in the sample space, which can make the current metric get its exact position in the sample space. A detailed introduction of RSD is presented in Ref. 6.

ADTCP is a friendly transport protocol and can achieve better performance. But the complexity of this protocol and the compatibility problem with TCP make it difficult to implement in real system.

3 Design an End-to-End TCP Enhanced Scheme for Ad Hoc Networks

Based on existing protocols and previous discussion, we benefit from certain advantages of ADTCP and attempt to design a simple and effective TCP end-to-end enhanced scheme in order to reduce the complexity of calculation.

3.1 Identifying the Network States

The new scheme can distinguish three network states, namely congestion (CON), channel error (CHERR) and route change (RTCHG). The identification for these network states based on end-to-end measurements is mainly achieved by TCP receiver, which successively dispatches the feedback on current network states and events to TCP sender via the ACK packets. The detailed process is depicted as follows.

1) Identifying Congestion

Firstly, we detect the potential CON state by using IDD (Inter-packet delay difference) as ADTCP. Based on the shortages of ADTCP mentioned in Section 2, we take measures in the new scheme: ① In order to improve the sensitivity of identifying congestion state, the sample space is based on the history records of a former and longer time, and the weight of each history record is uniform. In addition, the sample space is requested to rebuild after route change for eliminating the influence on the old route. ② IDD shows how acutely the buffer queue changes, and RTT is used to describe the length of the buffer queue. If IDD can cooperate with RTT to identify CON state, we may obtain better results. The sample space of RTT is maintained in the same way as the sample space of IDD. Namely, CON is estimated by comparing the current RTT with the value of the sample space. In this scheme, the necessary condition of CON is that both IDD and RTT are CON.

2) Identifying Non-congestion States

RTCHG: if the network state is not congestion, we next seek to detect whether the events of RTCHG just now happens or not by measuring "Hop Limit". As long as the network layer of the receiver submits "Hop Limit" of the arrived packet to the

transport layer, TCP receiver can speculate that the route has already been changed according to the change of “Hop Limit” of the arrived packet.

CHERR: if TCP detects packet loss when the events of CON and RTCHG do not happen, namely the current state is not congestion or route change, we may suppose that the current network state is channel error (CHERR).

3.2 Taking the Corresponding Measures

Data transfer proceeds with normal TCP operations until an interruption is triggered by either a re-transmission time-out or a third duplicate ACK. Consequently, the sender takes corresponding measures according to the estimation of current network state, as Fig. 1 shows.

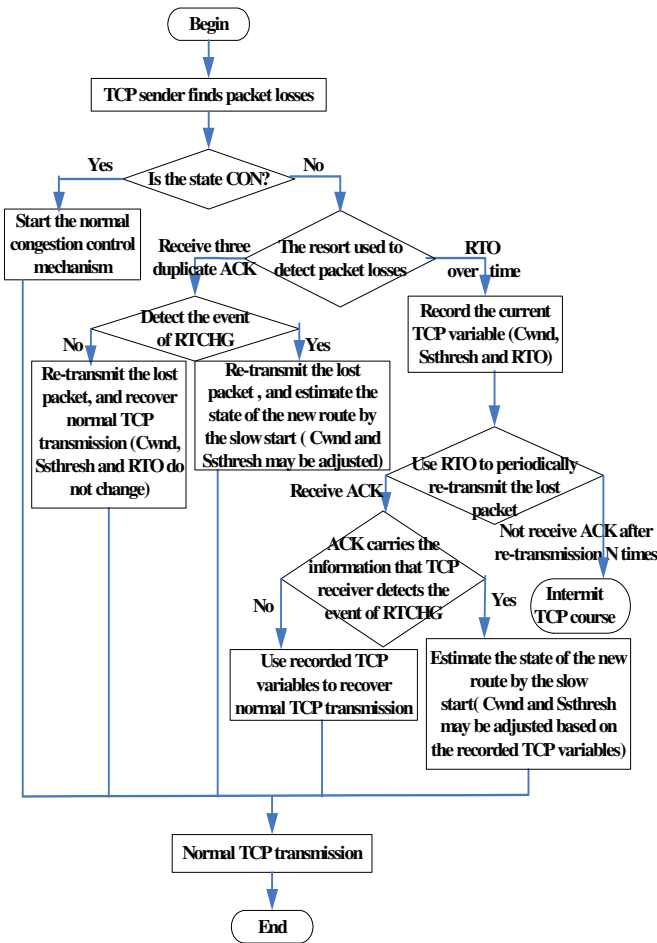


Fig. 1. The flow chart of corresponding measures taken by TCP receiver

- Congestion (CON): When network congestion occurs, ad hoc transport should adopt the same congestion control actions as conventional TCP3.
- Route change (RTCHG): In this case, the sender should estimate the bandwidth along the new route by setting its current sending window to the current slow start threshold, and initiating the congestion avoidance phase, and re-transmit the lost packet.
- Channel error (CHERR): When random packet loss occurs, without slowing down, the sender should re-transmit the lost packet.

4 Simulations of TCP Enhanced Scheme

This section evaluates the performance of the new TCP scheme (called New TCP in the simulation) relative to TCP NewReno, TCP ELFN and ADTCP. The simulation was done by network simulator NS-2.

In Fig. 2, it can be seen that NewReno’s performance is more sensitive to node mobility than New TCP. This is because as the mobility speed increases, network disconnection becomes more common. By correctly identifying non-congestion packet losses, New TCP is able to recover from such interruption quickly and achieve higher throughput.

Severe channel condition is added in simulation to illustrate the effects of channel error on TCP, where the channel error rate is set to be 5% in our simulation scenarios. In Fig. 3, the presence of channel error slightly increases the performance gap between New TCP and ELFN. The gap comes from the identification inaccuracy of New TCP.

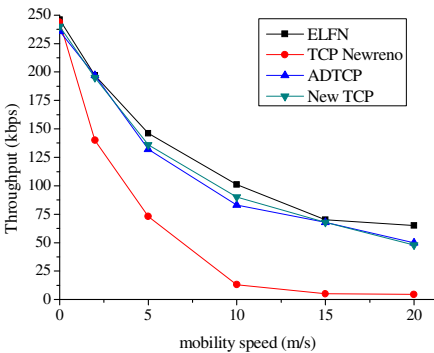


Fig. 2. Performance improvement of New TCP with mobility only

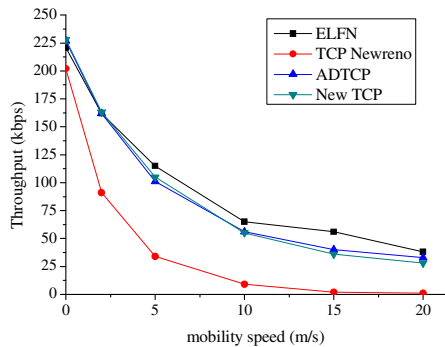


Fig. 3. Performance improvement of New TCP with mobility and 5% channel error

To introduce congestion, three competing UDP/CBR flows are run within the time intervals of [50, 250], [100, 200] and [130, 170], respectively. Each UDP flow transmits at 180 kbps. In Fig. 4 in which TCP flow competes with UDP/CBR flows, congestion becomes more frequent and the throughput gap between ELFN and New

TCP reduces. This shows that our identification algorithm detects network congestion state more accurately than non-congestion states.

Based on the simulation results analyzed above, we may conclude that New TCP could achieve better performance than TCP NewReno, and gain a performance close to TCP ELFN and ADTCP. However, New TCP belongs to end-to-end approaches, which is easy to implement and deploy, and requires no network support. Moreover, New TCP reduces the complexity of calculation in comparison with ADTCP.

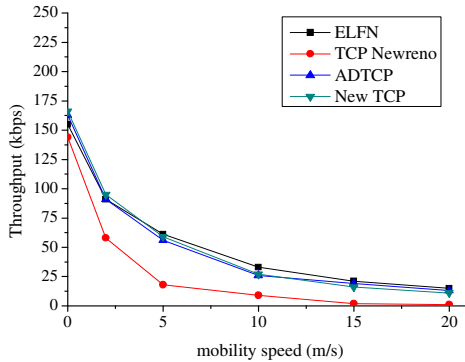


Fig. 4. Performance Improvement of New TCP with mobility, 5% channel error and 3 competing UDP/CBR flows

5 Analysis and Improvement of TCP Stability in Ad Hoc Networks

Due to the inherent problems of MAC protocol, routing protocol and TCP itself, the TCP flow is not stable in the wireless ad hoc networks based on IEEE 802.11. In this chapter, an enhanced algorithm with respect to IEEE 802.11 MAC protocol and DSR routing protocol is given. The simulation results show that the proposed algorithm can avoid the instability of TCP flow and increase its overall throughput in the ad hoc wireless networks.

5.1 IEEE 802.11 MAC and Dynamic Source Routing (DSR) Enhancements

There is also some theoretic analyses⁶ of the reason of TCP instability problem. We concluded that there are two main reasons of the TCP instability problem: the first one is that some nodes occupy the wireless channel so long that other nodes in the network can not access the channel within maximum RTS retry; the second reason is that if one node can not access the channel within the retry limit, then this node will send a route error message (but in fact this is false) to the source node thus triggering time-wasting route discovery process. During this process, the source node will stop sending data packets.

Based on the previous analysis, we propose enhancements to IEEE 802.11 MAC and DSR route protocol in this section.

5.1.1 Retry Limit in 802.11 MAC

We decide to modify the MAC protocol by adjusting the retry limit of 802.11 MAC which has been studied in Ref. 6. We have observed that if a response to a frame transmission is not received after the specified number of retries, the packet is dropped and link breakage reported. Also, the node drops all the packets destined to the same node in its IFQ. Under the above scenario, an improvement for MAC protocol would be to let the sender retransmit an increased number of times. Therefore, if we increase the retry limit, we expect to reduce the chances of a packet being dropped at the link layer. Furthermore, we expect a reduction in the number of route failures and oscillatory behavior. In many cases such an improvement will also increase overall throughput.

The current retry limit appears to be reasonable for wireless LANs with a base station, as every node is within the interference range of all other nodes. With a low retry limit, the node can detect link breakage early. If the retry limit is too high, it could end up wasting resources as a link failure due to mobility will take longer to be detected. Another drawback of high retry limit is that the transmission of other packets in the link layer queue (IFQ) is delayed due to the FIFO scheduling of IFQ. Overall, this results in longer end-to-end delays. In many cases in multi-hop wireless networks, the cost of unnecessary packet drop may far outweigh the drawbacks of high retry limits. This would be especially true in a low mobility situation. Thus, one of the possible ways to improve throughput of TCP based applications is by increasing the MAC protocol retry limit.

5.1.2 Dynamic Source Routing (DSR) Enhancements

Our proposed enhancement to DSR protocol is based on the work of Ref. 6. The algorithm is as follows: when the route layer of a node receives the link error message from the MAC layer which indicates that the link to the next hop is invalid, it does not send route error message immediately to the source node as DSR does. It sends a "Hello" message to the next hop. At the same time, a timer is triggered. If this node does not receive the reply of "Hello" message from the next hop until the timer expires, it then sends route error message to source node. Otherwise, if it receives the reply of "Hello" message from next hop before the timer expires, timer is canceled and no route error message will be sent to source node. Normal data transform is resumed. This process can efficiently determine if the link is really invalid thus avoid unnecessary route discovery process of DSR.

5.2 Simulations of IEEE 802.11 MAC and DSR Enhancements

Firstly we look into the effects of increasing the MAC protocol's retry limit. We increase the retry limit for short frames from 7 to 14 and for long frames from 4 to 10. We can see from Fig. 5 that modified MAC protocol can effectively alleviate the TCP instability problem. For window =32, there are only seven times that TCP throughput reaches zero. For window =8, no serious instability problem occurs at this level. We can also find from Table 2 that by using modified MAC protocol, the overall throughput is also improved.

Table 2. Summary of simulation parameters

	Maximum Window	Throughput(kbps)
IEEE 82.11 MAC with DSR	4	339.8
	8	262
	16	255.3
	32	232
Modified MAC with DSR	4	341.7
	8	349.4
	16	325.3
	32	303
IEEE 802.11 MAC with modified DSR	4	341.6
	8	328.3
	16	327.4
	32	332.5
Modified MAC with modified DSR	4	341.7
	8	349.4
	16	354.2
	32	354.6

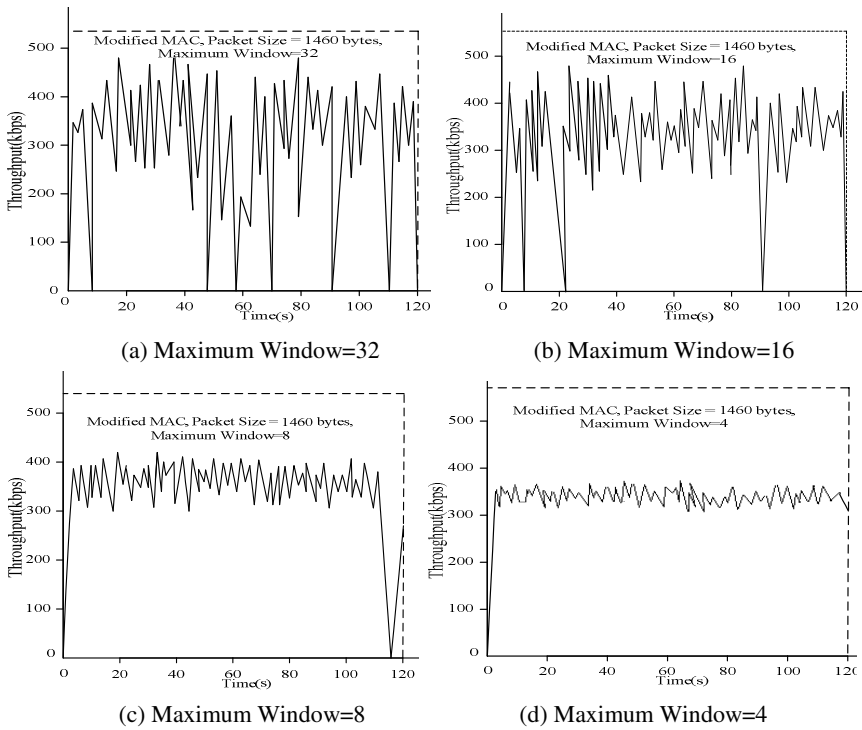


Fig. 5. The instability problem in the four-hop TCP connection (Modified MAC)

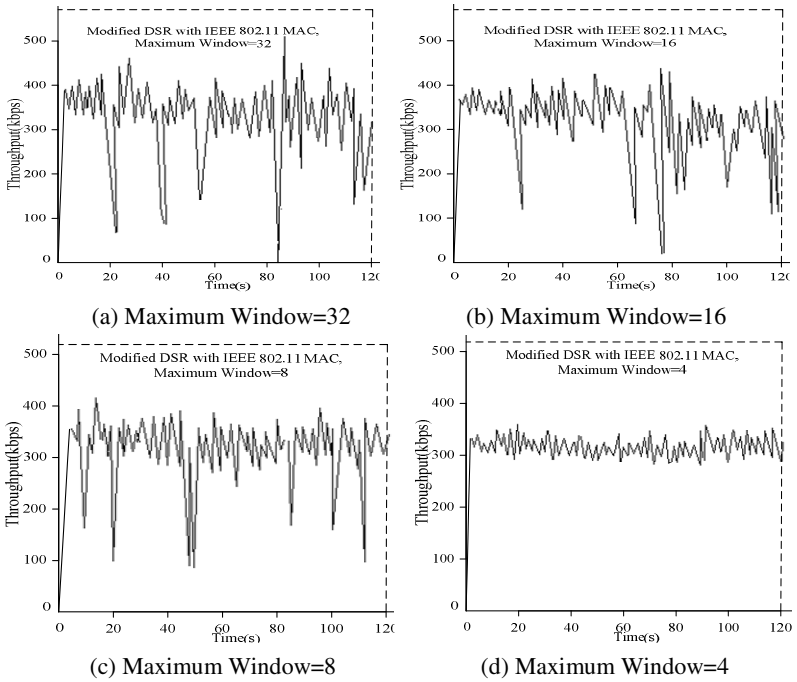


Fig. 6. The instability problem in the four-hop TCP connection (Modified DSR)

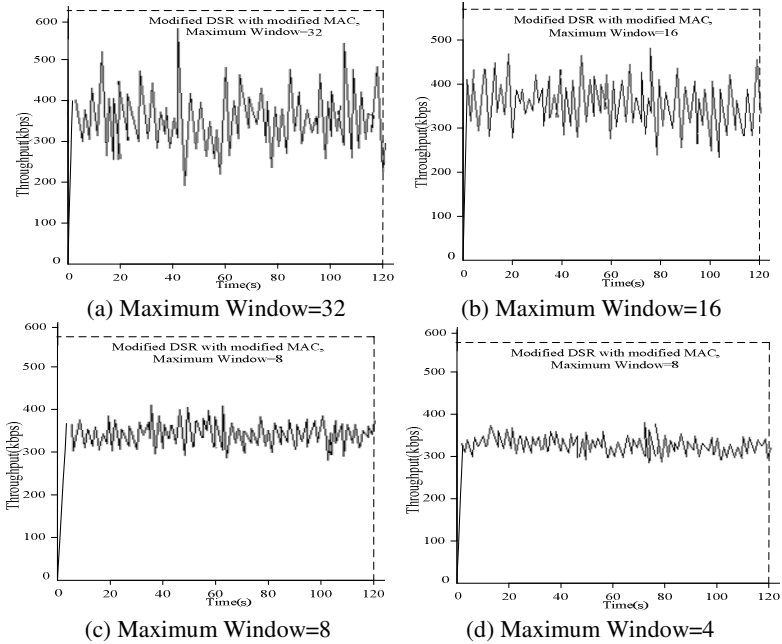


Fig. 7. The instability problem in the four-hop TCP connection (Modified DSR with modified MAC)

Then we use modified DSR with IEEE 802.11 MAC protocol. The simulation results are given in Fig. 6. It is obvious that modified DSR can effectively alleviate the TCP instability problem. We can also see from Table 2 that the overall throughput is also improved compared with "IEEE 802.11 MAC with DSR" scheme. From the simulation trace file, we found that modified DSR avoided several false route error messages by using "Hello" message exchanging. By avoiding false route error messages and unnecessary time-wasting route discovery procedure, TCP stability and overall throughput are improved.

Finally we use modified DSR with modified MAC. In the previous simulations, we have seen that MAC and DSR enhancement can alleviate the TCP instability problem respectively and they are in different network layers. We expect that combing these two enhancements, we can get better results. Simulation results show that our scheme is more efficient than using only MAC enhancement or DSR enhancement. This can be seen from both the Fig. 7 and the overall throughput comparison in Table 2. Fig. 7b shows the case with window = 16. No serious instability problem occurs at this level. But in previous simulations, there are still severe instability problem when window= 16. From Table 2 we can see that with each maximum window size setting, "Modified MAC with modified DSR" scheme has the best overall throughput.

6 Conclusions

In this paper our objective is to improve the performance of TCP in Ad hoc wireless networks. Towards this, a new End-to-End TCP enhanced scheme for ad hoc wireless networks is firstly presented. The simulation results with ns-2 show that the new scheme could achieve better TCP performance.

Then, we reveal the TCP instability problems in multi-hop wireless networks, which can degrade the TCP throughput performance seriously. So the IEEE 802.11 MAC and DSR enhancements are proposed to resolve this problem in this paper. Simulation results showed that the proposed scheme can not only avoid the TCP instability problem but also improve its overall throughput.

References

1. George, X., Polyzos, G.C.: TCP performance issues over Wireless Links. *IEEE Communications Magazine* 39(4), 52–58 (2001)
2. Mondal, S.A., Luqman, F.B.: Improving TCP performance over wired-wireless networks. *Computer Networks* 51(13), 3799–3811 (2007)
3. Huh, E.-N., Choo, H.: Performance enhancement of TCP in high-speed networks. *Information Sciences* 178(2), 352–362 (2008)
4. Majeed, A.: Analysis of TCP performance on multi-hop wireless networks: A cross layer approach. *Ad Hoc Networks* 12(3), 586–603 (2012)
5. Fu, Z., Greenstein, B., Lu, S.: Design and implementation of a TCP-Friendly transport protocol for ad hoc networks. In: *ICNP 2002, Paris, France*, pp. 216–225 (2002)
6. Xu, K., Gerla, M., Bae, S.: How Effective is the IEEE 802.11 RTS/CTS Handshake in Ad Hoc Networks. In: *IEEE GLOBECOM 2002*, pp. 72–76 (2002)

Part IV
Intelligent System Analysis and
Social Networks

SGR-StarCraft: Somatosensory Game Rehabilitation via StarCraft

Ching-Hsun Hsieh and Chia-Hui Wang

Department of Computer Science and Information Engineering,
Ming Chuan University, 5 De Ming Rd., Guei Shan District, Taoyuan County 333, Taiwan
{01366163@um, wangch@mail}.mcu.edu.tw

Abstract. Usually the patients who need rehabilitation have to go to hospital to complete the therapy. For most of patients, conventional rehabilitation is both time consuming and uninteresting. It also spends much medical and human resources in hospitals. Thanks to the advances in interactive technologies of somatosensory, those who have physical problems can have much more convenient and interesting ways for rehabilitation via somatosensory machine such as Kinect. In this paper, we use the StarCraft strategy game as a game rehabilitation example applied with Kinect somatosensory machine to design and develop an interactive platform which can help disabled people to complete therapy of rehabilitation at home. The proposed SGR-StarCraft¹ (Somatosensory Game Rehabilitation via StarCraft) system can correspondingly send different interactive commands to StarCraft game by checking the rehabilitation movement similarity between patient and rehabilitation professional. SGR-StarCraft also dynamically adjusts the levels of StarCraft playing difficulties to motivate patient to have willing to continue the game rehabilitation. Moreover, SGR-StarCraft can record the skeleton movement information during the game rehabilitation and medical professionals can use recorded data to advise patients to revise their rehabilitation movement in details for better therapies. In our experimental results, two improved methods applied from Euclidean distance and dynamic time warping (DTW) demonstrate their cost-effectiveness in calculating similarity scores of patients' rehabilitation movement for SGR-StarCraft.

Keywords: Kinect somatosensory machine, Rehabilitation, Euclidean distance, Dynamic time warping, StarCraft game.

1 Introduction

Rehabilitation therapy is an extended treatment for people with physical disabilities. With the evolution of time, professional rehabilitation is still a growing division. It is a very young medical science in the medical profession. It's how to get patients to be treated through these rehabilitation techniques and help patient use the good portion

¹ This work was partially supported by Ministry of Science and Technology, Project No. NSC 102-2221-E-130 -003.

of their body and function to the best, so that they can return home, back to the community, back to the work or live independently. Due to the meaningful rehabilitation therapies, it's often said that the medical science adds years for life, and the rehabilitation adds life for years.

However, the patients who need rehabilitation usually have to go to hospital to complete the therapy. Such conventional rehabilitation is both time consuming and uninteresting. Besides, conventional rehabilitation spends much medical and human resources in hospitals. Thanks to the advances in interactive technologies of somatosensory such as Kinect, those who have physical problems can have much more convenient and interesting ways for rehabilitation via the somatosensory machine. Kinect was originally designed to play video game via the player's body actions without controllers, because its cameras can perform 3D measurement and also render 3D depth images, and then the detected 3D depth information will be transformed into the skeleton tracking system.

The skeleton tracking system applied in Kinect is also helpful to track patient's rehabilitation movement for better therapy. While playing video game can be integrated to rehabilitation, the uninteresting rehabilitation may turn to be fun and cost-effective therapy. Since the patient's rehabilitation movement is needed to be scored to map to the different game challenging levels, the more similarity with standard rehabilitation movement from medical professional will have more chance to win the game to meet the goal of game rehabilitation. In this paper, we use well-known StarCraft strategy game [6] applied with Kinect somatosensory machine to design and develop an interactive game rehabilitation platform with movement scoring techniques to help disabled people to cost-effectively complete therapy at home with fun.

The remainder of this paper is organized as follows. In Section 2, we describe the related works of game rehabilitation including real-time strategy (RTS) games and Kinect skeleton tracking system. The applied techniques to integrate video games like StarCraft with Kinect's skeleton tracking system are introduced in Section 3. Issues and methods from scoring patient's rehabilitation movement and mapping to different game challenge levels are further described in Section 4. The performance results from proposed scheme for SGR-StarCraft (Somatosensory Game Rehabilitation via StarCraft) with different scoring algorithms are demonstrated in Section 5. Finally, we conclude this paper in Section 6.

2 Related Work

2.1 RTS Games

In recent years, RTS games such as Age of Empires [7], StarCraft series [6] are very prevalent and popular over Internet, and current game developers not only focus on the playing performance on screen presentation, but also emphasize the setting of artificial-intelligence (AI) opponents to enrich the game play. In the fierce competition market of RTS games, the computer AI technologies are no longer just a boost on intellectual strength of RTS games, also enable players to feel more

interesting, and feel like a real battle in reality. AI setting to RTS game is currently a very important issue. Usually, RTS games only provide a few play levels in default against computer AI. They cannot effectively cover the majority of today's game players with different ages. For example, game levels in StarCraft II contain very simple, easy, normal, hard, very difficult and crazy levels. When the entry-level players play the game against the chosen AI opponents from more difficult levels, AI opponents will usually have landslide victory; players may lose confidence and interest to continue to play RTS games. Therefore, how to find the ways for players to continue to play by simultaneously learning to match AI opponent with more difficult levels is a very interesting topic [8].

In general, most of RTS games generally include the roles of workers who are specialized to collect resources, soldiers for attack and various buildings. They perform continuous gathering resources, building bases and producing soldiers to attack enemies to win the game. In StarCraft II, crystalline mineral and gas are the resources to be collected by the workers to construct related buildings. The buildings can produce warriors and the warriors are used to attack opponent workers or warriors. The most fun part of the RST game is depending on different tactical to do a specific allocation of resources including produce workers, explore resources, allocate resources to produce the soldiers and further compose sufficient army. Therefore, the gathered army will attack and destroy the enemy to achieve the game victory.

2.2 Kinect Skeleton Tracking System

Kinect was originally initiated from Project Natal [1], Microsoft announced the official name Kinect from Project Natal in 2010. Kinect combines the terms of kinetics and connection with the slogan of "your body is the controller". The major difference with ordinary camera is that Kinect can perform 3D measurement via light coding theory [5]. Because the light coding technique just provides basic image depth data, Kinect system apply the detected 3D depth information and further transform into the skeleton tracking system. The Kinect system can simultaneously detect up to six people's movement, including two people's complete action information at the same time; each person can be recorded by a total of 20 sets of movement details [13] including the trunk, limbs and fingers to reach full sense of the whole body actions.

3 Proposed Somatosensory Game Rehabilitation via StarCraft (SRG-StarCraft)

3.1 StarCraft Map Editor

StarCraft II provides a map editor tool for players to customize their play games. Players can control a lot of detailed instructions such as the screen setting and the role modules, through the built-in scripting language. To design a game through StarCraft map editor, the main steps can be divided into following items: establishment of game

map, game initialization, creation of events and variables. These steps are described briefly as follows:

Establishment of game map:

To select the terrain object, players can draw their own game map through a variety of terrain objects.

Game initialization:

To press trigger button, player can generally set the starting position for the army, the game variables, events and etc.

Creation of events and variables:

An event object includes three elements of events, conditions and actions. An event element indicates this event is established under what kind of condition. Usually an event is set on map initialization and then this event can be valid as game initialization. The element "conditions" of event object indicates which action will be executed on some condition. These conditions include a particular variable is less than, equal to, and greater than a number, and etc. The element "action" of the event object will trigger and perform some action while a condition is met. These actions include game parameter changes, to enable AI opponent's actions (such as to produce units and attack) and so on.

3.2 SGR-StarCraft: Somatosensory Game Rehabilitation via StarCraft

In this RTS game of StarCraft II, player can obviously find basic strategy of AI opponents, namely the production of soldiers, production of workers, construction of buildings, and then attack. In [8], they mentioned that the most impact of the outcome of StarCraft game is during the both armies at war. Moreover, the most impact factors that influence the result of armies at war is the production of soldiers, so to control the production of soldiers will become the main cause of game victory.

However, the soldiers are divided into three parts, namely harassment unit, auxiliary unit and the main power unit. Harassment unit mainly pre-harassing enemy workers to affect the speed of access resources, but the combat capability is more vulnerable during the middle or the late of game, and then their usage will begin to decline.

Auxiliary unit itself is unlikely to have an offensive force, but it can strengthen the combat capability partners, it has a certain influence in late of the game. The main power unit is to play a major role during attack, and can join with and auxiliary units to strengthen their own combat capabilities, it can mainly influence the outcome of the game.

Therefore, SGR-StarCraft game rehabilitation therapy can use Kinect to score the similarity with correct rehabilitation movement, different scores such as the different levels of the similarity is very good, ordinary and bad, and then they will trigger the respective soldiers with different impact to the game victory, say main power unit, auxiliary unit and the harassment unit. Then, during the game playing time, SGR-StarCraft will change the proportion of these three different units of soldiers to make harassment unit won't be produced more than ever even if the lower score of

rehabilitation movement, since harassment unit has greater influence in the earlier stage of the game playing time.

4 Rehabilitation Movement Scoring Methods for SGR-StarCraft

4.1 Improved Methods from Euclidean Distance

The human skeleton information can be continuously captured by the Kinect system for SGR-StarCraft. The captured information is skeleton points of three-dimensional space, so the rehabilitation movement can be seen as an array of three-dimensional vectors in 3D space. Therefore, when comparing patient's rehabilitation movements with rehabilitation professional's for similarity and correctness in rehabilitation therapy, two different captured arrays of 3D vectors from patient and professional respectively can be used to obtain the rehabilitation score. Euclidean distance calculation is the common method to apply to check the similarity and correctness between two arrays of 3D points. While the distance value is high, we can say the similarity is low, and vice versa. However, the lengths of these two arrays are not the same usually since rehabilitation durations from patient and professional may not be exactly the same. Original Euclidean distance requires that the lengths two compared arrays are needed to be the same, so we propose two improved methods of Euclidean distance for rehabilitation movement scoring.

Averaging Compute-Euclidean Distance (ACED)

In ACED method, the longer array of 3D points will be instantly averaged to the same length as the shorter array of 3D points. The details of ACED method are shown as follows:

```

<Input parameters>
A={a1, a2, a3, ..., an}
B={b1, b2, b3, ..., bm} /* A and B are arrays of 3D points with length n m */
<Longer array transformation>
if(n>m) /* when length of A is greater than B */
  /* longer array A is averaged to shorten its length to m */
  A={ (a1+...+an-m+1)/(n-m+1), (a2+...+an-m+2)/(n-m+1), (a3+...+an-m+3)/(n-m+1), ...,
    (am+...+an)/(n-m +1) };
else if(m>n) /* when length of B is greater than A */

```

Fig. 1. ACED Algorithm

Through this method two arrays of 3D points can be converted into two arrays with the same length in a short time. Therefore, the improved Euclidean distance can be applied further to check similarity of rehabilitation movement between patient and professional. However, the longer array will be shortened to the same length as shorter array. Though the computation complexity of ACED is low, the cost is the

scoring accuracy of similarity between rehabilitation movements because of the reduced length of one array of 3D points. Therefore, the following proposed method applies the interpolation for shorter array to increase array length to prevent the cost of scoring accuracy.

Repeated Interpolation-Euclidean Distance (RIED)

The proposed RIED method is composed of two parts: 1) when the array length of 3D points from standard movement of rehabilitation professional is greater than the rehabilitation movement of patient, RIED will be randomly choose two adjacent points in shorter array, average them and insert the averaged point among these two points. Then, this interpolation step can be continuously repeated to increase the shorter array length until the array length equals to longer array; 2) when the array length of 3D points from patient's rehabilitation movement is greater than standard movement, we convert these two arrays into arrays with the same length like ACED shown in Figure 1. Through this RIED method, without sacrificing and 3D points of standard movement, interpolation and instantly averaging are applied for shorter and longer array length respectively from patient's rehabilitation movement. Then, the Euclidean distance can be easily calculated on two arrays with the same array length to evaluate the degree of rehabilitation movement similarity with each other. In Figure 2, the RIED method is described in details.

```

<Input parameters>
A={a1, a2, a3, ..., an}; /* array A is the standard movement */
B={b1, b2, b3, ..., bm}; /* arrays of A and B with length n and m */
int ld = n-m;
<array transformation>
while(ld>=1){ /* length of A is greater than B */
    int rN=random(0~m-1); /* pick a random number from 0 to m-1 */
    if(ip[rN]!=NULL){
        ip[rN]=(B[rN]+B[rN+1])/2; /* generate interpolation points */
        ld--;
    }
}
for (i:0 to n-1 j:0 i++,j++){ /*insert interpolation points to B*/
    B_tmp[i]=B[i]; /* temporary array for interpolation */
    if(ip[j]!=NULL){
        B_tmp[i+1]=ip[j];
        i++;
    }
}
}

```

Fig. 2. RIED Algorithm

4.2 Dynamic Time Warping

Dynamic time warping algorithm was first developed in 1978 by Hiroaki Sakoe and Seibi Chiba [9], the DTW algorithm is usually used for voice recognition; its basic definition is mainly used to measure the similarity between two data sequences in which the time or speed may be different. Suppose there are two sets of vectors called X and Y , their lengths are m and n respectively, the purpose of DTW is to find a path P which goes through p_1, p_2, \dots , and p_K . The p_k is formed by (x_i, y_j) , i represents the i -th point in vector X and j represents the j -th point in vector Y . Besides, DTW also follows the rules of endpoint and local relations:

Endpoint relation: DTW calculation is required to satisfy the starting point p_1 equals to (x_1, y_1) and end point p_K equals to (x_m, y_n) .

Local relation: DTW calculation is satisfy when $p_k=(x_i,y_j)$, the nearby points must be $p_{k+i}=(x_{i'},y_{j'})$, $i \leq i' \leq i+1$ $j \leq j' \leq j+1$ Which $Dist(P) = \sum_{k=1}^K Dist(p_{ki}, p_{kj})$, p_{ki} is the i -th

point in vector X and p_{kj} is the j -th point in vector Y , k represents the current count of points in the shortest path. Where $Dist(p_{ki}, p_{kj})$ is any way to calculate the distance between these two points. In this paper, the Euclidean distance is applied. The $Dist(P)$ is a value of the shortest path P and this value is calculated by DTW algorithm.

A recursive formula from Equation (1) to calculate DTW values is expressed as follows:

$$D(i, j) = Dist(i, j) + \min[D(i-1, j), D(i, j-1), D(i-1, j-1)] \tag{1}$$

The $D(i, j)$ of Equation (1) is the DTW distance between $X(1 \sim i)$ and $Y(1 \sim j)$, $Dist(i, j)$ indicates the increased cost through this shortest path from $(1, 1)$ to (i, j) . For example, we assume the values of vector A and B are shown as follows:

Table 1. Vector A

	a1	a2	a3	a4
x-axis	2	5	3	7
y-axis	3	4	2	5

Table 2. Vector B

	b1	b2	b3
x-axis	6	4	2
y-axis	4	2	2

First, we will calculate the cost distance between any two points of vectors A and B , The following table is represented with a two-dimensional array:

Table 3. DTW Distance Cost Table

	<i>a1</i>	<i>a2</i>	<i>a3</i>	<i>a4</i>
<i>b1</i>	$\sqrt{(2-6)^2 + (3-4)^2} = 4.12$	1	3.61	1.414
<i>b2</i>	2.24	2.24	1	4.24
<i>b3</i>	1	3.61	1	5.83

Then according to the distance cost table, we calculate the similarity cost which must be spent through all the points in the path and they are shown in Table 4:

Table 4. DTW Similarity Cost Table

<i>C</i>	1	2	3	4
1	(<i>a1</i> , <i>b1</i>) = <i>C</i> 11	(<i>a2</i> , <i>b1</i>)+ <i>C</i> 11 = <i>C</i> 21	(<i>a3</i> , <i>b1</i>)+ <i>C</i> 21 = <i>C</i> 31	(<i>a4</i> , <i>b1</i>)+ <i>C</i> 31 = <i>C</i> 41
2	(<i>a1</i> , <i>b2</i>)+ <i>C</i> 11 = <i>C</i> 12	(<i>a2</i> , <i>b2</i>)+ Min(<i>C</i> 11, <i>C</i> 21, <i>C</i> 12) = <i>C</i> 22	(<i>a3</i> , <i>b2</i>)+ Min(<i>C</i> 21, <i>C</i> 31, <i>C</i> 22) = <i>C</i> 32	(<i>a4</i> , <i>b2</i>)+ Min(<i>C</i> 31, <i>C</i> 41, <i>C</i> 32) = <i>C</i> 42
3	(<i>a1</i> , <i>b3</i>)+ <i>C</i> 12 = <i>C</i> 13	(<i>a2</i> , <i>b3</i>)+ Min(<i>C</i> 12, <i>C</i> 22, <i>C</i> 13) = <i>C</i> 23	(<i>a3</i> , <i>b3</i>)+ Min(<i>C</i> 22, <i>C</i> 32, <i>C</i> 23) = <i>C</i> 33	(<i>a4</i> , <i>b3</i>)+ Min(<i>C</i> 32, <i>C</i> 42, <i>C</i> 33) = <i>C</i> 43

The value of *C*43 is calculated from two vectors *A* and *B* by DTW algorithm, the smaller value of *C*43 indicates more similar between vectors *A* and *B*. [10], [11], [12] and others applied DTW to realize the movement similarity checking from Kinect skeleton information. The following section 5 will present the experiments and performance results from ACED, RIED and DTW algorithms for scoring rehabilitation movements in proposed SGR-StarCraft system.

4.3 Normalization of Scoring

Since the above-mentioned scoring algorithms preserve different ranges in their scoring systems, these different values need to be normalized to conventional scoring range like 0 to 100, higher value indicates higher similarity in rehabilitation movement with standard movement. Then the normalized score can be transformed to different challenging levels in StarCraft II as mentioned in Section 3.

$$E_s = \sqrt{\frac{(Max(|X_s - X_{max}|, |X_s - X_{min}|))^2 + (Max(|Y_s - Y_{max}|, |Y_s - Y_{min}|))^2}{(Max(|Z_s - Z_{max}|, |Z_s - Z_{min}|))^2}} \quad (2)$$

The proposed ACED and RIED are improved from Euclidean distance, so the maximum distance value E_s from an array of standard 3D movement points (i.e. X_s, Y_s, Z_s) can be obtained by Equation (2). ($X_{min}, Y_{min}, Z_{min}$) and ($X_{max}, Y_{max}, Z_{max}$) indicate the minimum and maximum values of 3 different axis in Kinect 3D skeleton system.

The largest value (i.e. worst score) for similarity with standard movement is E_s in Equation (2).

$$E_r = \sqrt{(X_s - X_r)^2 + (Y_s - Y_r)^2 + (Z_s - Z_r)^2} \quad (3)$$

Then, patient's rehabilitation movement will have 3D points like (X_r, Y_r, Z_r) and the corresponding Euclidean distance score E_r is shown in Equation (3). Finally the normalized $Score_{ED}$ ranged from 0 to 100 can be easily found in Equation (4).

$$Score_{ED} = 100 - E_r / E_s * 100 \quad (4)$$

In normalization of DTW scoring, we define an array whose length equals to the length of standard movement array. The defined array ranges from $(X_{min}, Y_{min}, Z_{min})$ to $(X_{max}, Y_{max}, Z_{max})$ and it will also have largest value (worst score D_s) of DTW scoring with 3D points from standard movement. Then, the DTW score of similarity between standard movement and patient's movement is D_r . Similar to Equation (4), we can have Equation (5) to normalize the DTW score D_r to $Score_{DTW}$.

$$Score_{DTW} = 100 - D_r / D_s * 100 \quad (5)$$

5 Experiments and Performance Results from SGR-StarCraft

In this section, we will validate the scoring accuracy of patient's rehabilitation movement on proposed SGR-StarCraft game rehabilitation system. As shown in Figure 3, the shoulder rehabilitation movement for rheumatoid arthritis is conducted on our experiments. The experiments and performance results are described as follows.

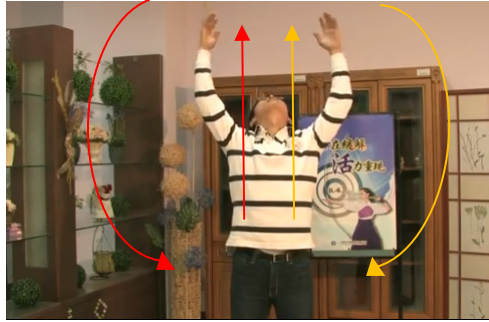


Fig. 3. Shoulder Rehabilitation Movement

5.1 Experiments for Patient's Different Rehabilitation Movements

The shoulder rehabilitation movement for rheumatoid arthritis is forced patient to hand up and extend as far as to the left and right as possible and then return to the original, as illustrated in Figure 3. We conduct 4 kinds of different shoulder rehabilitation movements from patients to validate our proposed scoring methods for SGR-StarCraft. The first one is very similar to standard movement from rehabilitation professional. The second one is similar to the first one, but faster movement than the

standard movement. The third one is similar to the first one, but slower movement than the standard movement. The last one is irrelevant movement to shoulder rehabilitation and patient just moves randomly. The 3D traces of standard movement and other 4 different kinds of above-mentioned patient’s rehabilitation movements are all illustrated in Figure 4.

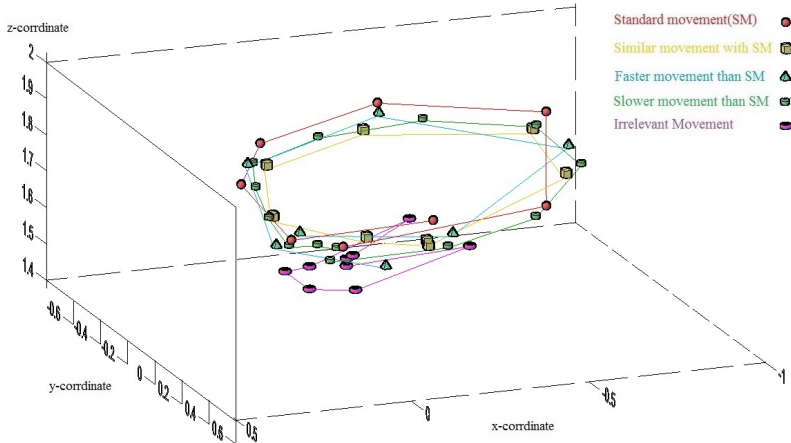


Fig. 4. 3D Traces of Different Rehabilitation Movements

5.2 Scoring Accuracy from ACED, RIED and DTW for SGR-StarCraft

The results from ACED, RIED and DTW scoring methods applied on 4 different patient’s movement are illustrated on Figure 5. Though ACED and RIED are improved methods from Euclidean distance with less computation complexity than DTW, DTW can provide more accuracy in rehabilitation movement scoring than others for SGR-StarCraft system. In the light-weight scoring methods of proposed ACED and RIED, due to the applied averaging to remove the 3D points in ACED method, the irrelevant movement unbelievably reaches higher score than faster and slower movements. Therefore, RIED can provide better scoring accuracy than ACED since RIED applies interpolation techniques without reducing 3D points in movement data set.

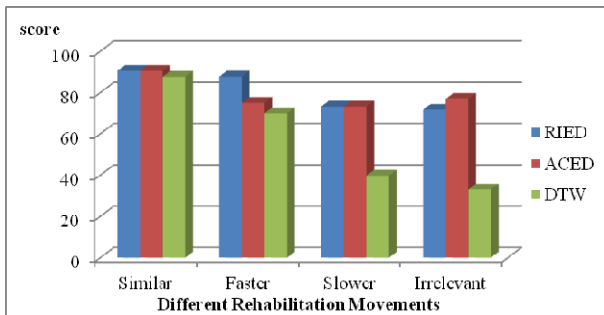


Fig. 5. Scoring Results from ACED, RIED and DTW

6 Conclusion and Future Work

In this paper, a somatosensory game rehabilitation system called SGR-StarCraft is proposed. SGR-StarCraft applied the RTS game of StarCraft II's map-editor tool to customize this RTS game feasible for rehabilitation therapy with game playing fun. Two improved algorithms (i.e. ACED and RIED) from Euclidean distance and well-known DTW are applied for scoring patient's rehabilitation movement compared with standard movement of rehabilitation professional. The performance results from our preliminary experiments on shoulder rehabilitation movement for rheumatoid arthritis demonstrate the cost-effectiveness in proposed scoring algorithms for SGR-StarCraft system.

Since the rehabilitation therapy is probably more complex than the scoring algorithms applied in proposed SGR-StarCraft, in the near future, not only more rehabilitation movements will be examined as many as possible on SGR-StarCraft in near future, but also we would like to consult with professional doctors and rehabilitation patients in hospital to discuss further improvements on proposed SGR-StarCraft system.

References

- [1] Kinect introduction, <http://zh.wikipedia.org/wiki/Kinect>
- [2] Fan, Y.-C.: The Performance and Motion Analysis of Virtual Reality Combined With Motion-Sensing Technology in Shoulder Joint Rehabilitation System (2013), <http://ir.lib.ncu.edu.tw/handle/987654321/61605>
- [3] Loongo somatosensory, <http://longgood.com.tw/>
- [4] Hong Kong Polytechnic University somatosensory game "step on cockroaches" Helping rehabilitation, http://www.uonline.nccu.edu.tw/index_content.asp?sn=6&an=14227
- [5] Light Coding, <http://book.51cto.com/art/201211/368691.htm>
- [6] StarCraft introduction, <http://tw.battle.net/sc2/zh/>
- [7] Age of Empires introduction, [http://en.wikipedia.org/wiki/Age_of_Empires_\(video_game\)](http://en.wikipedia.org/wiki/Age_of_Empires_(video_game))
- [8] Chang, S.-H., Yang, N.-Y.: A Study of Dynamic Challenging Level Adapter for Real-time Strategy Games. In: IEEE 15th International Conference on Computational Science and Engineering (2012)
- [9] Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics, Speech and Signal Processing 26(1), 43–49 (1978)
- [10] Kopaničáková, A., Virčíková, M.: Gesture Recognition using DTW and its Application Potential in Human-Centered Robotics. Robotics research paper (2013)
- [11] Su, C.-J., Huang, J.-Y., Huang, S.-F.: Ensuring Home-based Rehabilitation Exercise by Using Kinect and Fuzzified Dynamic Time Warping Algorithm. In: Proceedings of the Asia Pacific Industrial Engineering & Management Systems Conference (2012)
- [12] Jiang, H., Jie, X.: Kinect-based Rehabilitation Training Assistant System Research and Implementation. In: 2013 International Conference on Software Engineering and Computer Science. Atlantis Press (2013)
- [13] http://tw-hkt.blogspot.tw/2012/02/kinect-for-windows-sdk-v1_1798.html

Discovering Sentiment of Social Messages by Mining Message Correlations

Hsin-Chang Yang¹, Chung-Hong Lee², Chun-Yen Wu³, and Yu-Chian Huang³

¹ Dept. Information Management, National University of Kaohsiung, Taiwan
yanghc@nuk.edu.tw

² Dept. Electrical Engineering, National Kaohsiung University of Applied Sciences,
Kaohsiung, Taiwan

leechung@mail.ee.kuas.edu.tw

³ Institute of Information Management, National University of Kaohsiung, Taiwan

Abstract. With explosive growth of the Internet, the amount of information in text form is growing rapidly and the demand for data analysis is also increases. We can perform sentiment analysis on a large set of text messages to discover valuable knowledge and obtain enormous benefits in national security, business, politics, economics, etc. However, text messages from the social networks are rather different from those of traditional text documents. Therefore, it is difficult but essential to develop an effective method of sentiment exploration in social networks. In this paper we first applied a neural network model, namely the self-organizing maps, to cluster similar messages and sentiment keywords, respectively. We then developed an association discovery process to find the associations between a message and some sentiment keywords. The sentiment of a message is then determined according to such associations. We performed experiments on Twitter messages and obtained promising results.

Keywords: Sentiment Analysis, Social Network Analysis, Text Mining, Self-Organizing Map.

1 Introduction

Recently, social network services and sites emerged rapidly. According to a report by comSCORE in 2011 [1], social networking had surpassed search engines and become the most popular online activity worldwide. It accounted for nearly 1 in every 5 minutes spent online in October 2011, and reaches 82 percent of the worlds Internet population, representing 1.2 billion users around the globe. Meanwhile, Nielsen [2] also reported that Facebook is the top U.S. web brand in terms of time spent, as some 17 percent of time spent online via personal computer is on Facebook in 2012. These facts show that social network services have become one of the major communication channels among people.

The social network services are multi-faceted, including message posting, web site hosting, blogging, profile management, audio/video uploading, and apps,

etc. It is a common image that social network users often playing games and apps. However, according to a survey on real usage of Facebook [3], Facebook users spent 27% of time on Newsfeed (rank 1 in 5 task categories) while spent only 10% on apps. In fact, message posting seems to be the major activity between users in social network services. A famous example is Twitter, which provides microblogging service, reaches 1 in 10 Internet users worldwide to rank among the top social networks, and posted an impressive growth rate of 59% over 2011 [1]. Furthermore, the use of Twitter around the world has not been limited to interpersonal communication among friends. It was widely used as a central means of communication during events of worldwide and national significance. An example is that the tweets (a message sent using Twitter) became the major news source when the earthquake and tsunami struck Japan in 2011 and caused telephone networks to fail. Therefore, sending messages through social network services have been one of the most important channels for policy promotion, commercial advertising, and personal communications.

There are various types of messages used in social network services, such as texts, images, and videos. However, texts are always the most timely and widely used media used in social networks. As an example, Twitter sends out more than 500 millions messages per day [4]. Therefore, text social messages should be able to provide various kinds of knowledge that can be used in multiple purposes, such as advertising, security, and forecasting, etc. The analysis and mining of textual social messages thus received much attention recently. Unlike traditional social network analysis or mining techniques which focus on deriving properties from connections between users, mining textual social messages is considerably difficult if we were allowed to use only the message contents. Actually, this task resembles to the *text mining* that were widely investigated in the past decades. However, mining social messages is comparatively difficult for they have some unique properties:

- The lengths of social messages are generally bi-polar. As an extreme example, some messages posted in Facebook may contain only single word. On the other hand, a message may also consist of several thousands of words. For example, a user may copy the content of a survey article and post it on his blog.
- The social messages may contains words that carry little meaning in ungrammatical manner. Expression symbols are also widely used.
- The amount of social messages are tremendous compared to general media. For example, on August 3, 2013 in Japan, people watched an airing of Castle in the Sky, and at one moment they took to Twitter so much that it hit a one-second peak of 143,199 Tweets per second [4].

Sentiment analysis or opinion mining is a popular topic for social network applications since it is crucial for knowing the emotions of people for commercial and political purposes. For example, if we can discover the opinions of one person on some political issues, we can send him proper promotional materials to enhance or change his political preference. Another example is to detect the emotional state of some user to prevent possible damage on his mental or

physical status. Enterprises can also collect opinions on their products for further advertising and improvement. Therefore, sentiment analysis has attracted lots of attention in social network analysis research. However, mining sentiment of social messages is not easy due to above-mentioned characteristics of social messages. Therefore, traditional text mining approaches may not meet the need of social message mining tasks such as sentiment analysis. Customized feature extraction and mining process should be applied to conquer such challenges.

In this work, we will propose an approach to discover the sentiments or opinions underlying social messages. First, a feature extraction process suitable for social messages is proposed. Second, a message mining process based on self-organizing map algorithm [5, 6] is used to discover the sentiment or opinion underlying a message. We conducted experiments using messages collecting from Twitter and demonstrated the plausibility and effectiveness of our approach.

The remaining text is divided into following sections. We first review some articles that are related to this research in Sec. 2. After the literature review, we will discuss the proposed method in detail. First, the preprocessing and clustering of social messages using self-organizing map (SOM) algorithm is described in Sec. 3. We then show how to discover the sentiment of messages according to the clustering result in Sec. 4. In Sec. 5 we will demonstrate the experimental result. Finally, we give conclusions in the last section.

2 Related Work

We discuss some works related to our research here. One of the main task of sentiment analysis is to classify a text segment into some polarities such as positive, neutral, and negative. On this regard, the definition of polarity scales is crucial in sentiment analysis. Two of the pioneer works appeared in Turney [7] and Pang et al. [8]. Pang and Lee [9] further improved their scheme by allowing multi-scale emotional classes besides basic polarities. Such multi-scale polarities were also adopted by Snyder and Barzilay [10] in representing the users' opinions on different aspects of restaurants. Thelwall et al. [11] further extended the scale to be continuous to evaluate the sentiment of short texts. Kim and Hovy [12] adopted a different scheme for emotional polarities. They tried to extract the sentences for pros and cons from some product reviews.

3 Message Clustering

In this work, we will first cluster messages to reveal their relationships. First we should encode the messages into a set of vectors according to a vocabulary of polarized words. This vocabulary contains thousands of words that have been assigned to their sentimental polarity. A message M_i is then encoded into a vector $\mathbf{M}_i = \{m_{ij} | 1 \leq j \leq N\}$, where N denotes the size of the vocabulary $V = \{v_j\}$, using

$$m_{ij} = P(v_j), \quad (1)$$

where v_j is the j -th keyword in V and $P(v_j)$ is its polarity score. Generally, $P(v_j)$ has bi-polar value of either 1 for positive polarity or -1 for negative polarity. We then constructed a self-organizing map to cluster the message vectors. We should obtain a map \mathcal{M}_M which was trained by the message vectors \mathbf{M}_i . A labeling process was then applied on the map to label each message M_i on one of the neurons in the map. After the labeling process, we may obtain a clustering of these messages, known as the message cluster map (MCM). In this map, each neuron represents a cluster which consists of a set of messages containing similar words.

We also constructed another SOM to cluster the keywords in the vocabulary. For keyword $v_j \in V$, we may encode it by a vector $\mathbf{v}_j = \{\nu_{jl} | 1 \leq l \leq L\}$ using the following equation:

$$\nu_{jl} = w_{lj}, \quad (2)$$

where L is the number of neurons in \mathcal{M}_M and w_{lj} is the j -th element of the l -th neuron. These vectors are trained using SOM algorithm into a map \mathcal{M}_V . We then labeled each keyword v_j to one of the neurons and obtained the keyword cluster map (KCM). In this map, each neuron represents a cluster which consists of a set of keywords often co-occurred in the same set of messages.

4 Sentiment Analysis Approach

We applied a two-step scheme to discover the sentiment of a message. First, the polarity of a keyword cluster in the KCM is calculated. We then discover the associations between a message cluster in MCM to some keyword cluster in KCM. A message's sentiment polarity can then be obtained through such associations.

First, the polarity score of a keyword cluster K_j , $P(K_j)$, can be calculated by averaging the polarity scores of its constituent keywords as follow:

$$P(K_j) = \frac{\sum_{v_l \in K_j} P_0(v_l)}{|K_j|}, \quad (3)$$

where $P_0(v_l)$ denotes the polarity score of keyword v_l which is a constituent keyword of cluster K_j . It is obvious that if cluster K_j contains many positive keywords, $P(K_j)$ will be large, or vice versa. Through this approach, we can detect the sentiment polarity of every keyword clusters.

On the second step, we will try to discover the associations between each message cluster and some keyword cluster. A message cluster is considered to be relevant to a keyword cluster if the theme of the message cluster can be reflected by the keywords in the keyword cluster. Here, the theme of a message cluster is represented by a set of words that reveal the main idea of those messages belonging to this cluster. For example, if the theme of a message cluster is 'multimedia+Web+format' and these thematic keywords all appear in some keyword cluster, we can then consider them to be related. According to such

comprehension, we define the relevance between a message cluster C_i and a keyword cluster K_j by

$$R(K_j, C_i) = \frac{1}{N_c(C_i)} \sum_{C \in N_c(C_i)} \frac{\mathbf{w}_{K_j} \cdot \mathbf{w}_C}{\|\mathbf{w}_{K_j}\| \|\mathbf{w}_C\|}, \quad (4)$$

where $N_c(C_i)$ is the neighborhood of C_i in the MCM, \mathbf{w}_{K_j} is the synaptic weight vector of the neuron associated with K_j in the KCM, and \mathbf{w}_C is the synaptic weight vector of the neuron associated with C in the MCM. The keyword cluster K_j will be associated with the message cluster C_i if

$$i = \underset{k}{\operatorname{argmax}} R(K_j, C_k). \quad (5)$$

The sentiment polarity of an incoming message M_I can then be determined by first labeling it to some cluster C_I in the MCM according to the following equation:

$$I = \underset{k}{\operatorname{argmin}} \|\mathbf{M}_I - \mathbf{w}_{C_k}\|, \quad (6)$$

where \mathbf{M}_I denotes the message vector of M_I as described in Eq. 1. The sentiment polarity score of M_I can then be determined as $P(K_j)$, where K_j is the associated cluster of C_I .

5 Experimental Result

We performed experiments on a corpus of social messages collected from Twitter between January 2012 and March 2012. A total of about one hundred millions messages, also known as Tweets, were collected. For preliminary evaluations of our approach, we selected 7000 Tweets randomly from the corpus as the training set. We also selected another 3000 Tweets as the testing set. The minimum length of these Tweets is set to 3 to avoid extreme short messages that may contain no sentimental words. We discarded non-English words and embedded multimedia objects in these Tweets that are irrelevant to later processing.

These Tweets were transformed into vectors and clustered as described in Sec. 3. Generally, the vocabulary of the training set is constructed by collecting all words occurred in the set. However, we did not adopt this approach but use a standard opinion word vocabulary¹ collected in [13] since those words have been classified according to their polarities. The vocabulary in experiments contains words that appear in both message set and the opinion word vocabulary. The training parameters of the SOM used in the experiments are shown in Table 1. These parameters were determined experimentally to achieve best results. We tried different learning rates from 0.1 to 1.0 and maximum training epoch counts from 200 to 700. We obtained the message cluster map as well as the keyword cluster map of the training set after the SOM training.

¹ Downloaded from <http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>

Table 1. The training parameters of the SOM and statistics of data

Parameters	MCM	KCM
Size of vocabulary/Dimension of vectors	1847	100
Number of training data	7000	1847
Size of map	10 × 10	10 × 10
Learning rate	0.7	0.7
Maximum training epoch count	400	600

The polarity score of each keyword cluster in the KCM was then calculated by Eq. 3. Fig. 3 depicts the polarity scores of all keyword clusters. A score of 1 means that all keywords in this cluster have positive polarity. Meanwhile, score -1 means that all keywords have negative polarity. The dashed circles depicted clusters with neutral polarity. We can observe that most clusters have similar polarity as their neighbor clusters.

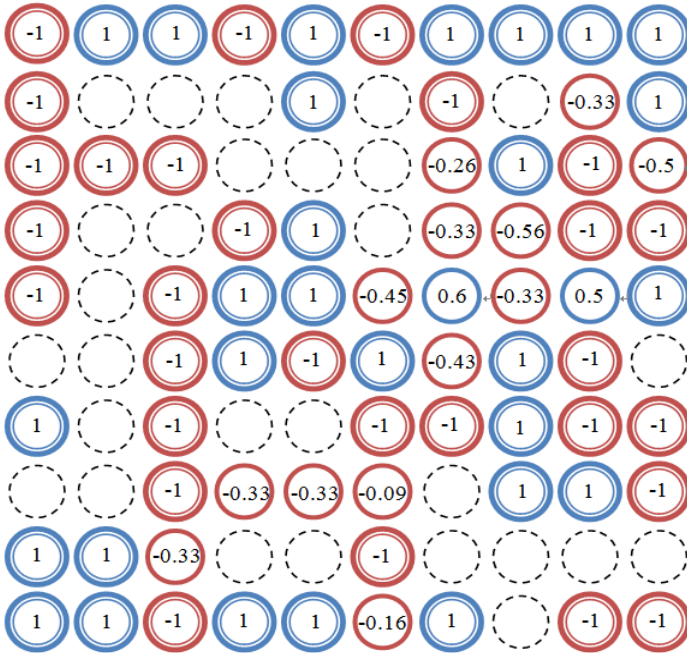


Fig. 1. The polarity distribution of keyword clusters

The associations across message cluster and keyword clusters were obtained as described in Sec. 4. Fig. 2 depicts the discovered associations between the MCM and KCM. Here C_i and K_j denote the indices of message clusters and keyword clusters in the MCM and KCM, respectively. Their relevance scores are determined by Eq. 4 and depicted as $R(J_j, C_i)$ in this figure.

C_i	K_j	$R(K_j, C_i)$	C_i	K_j	$R(K_j, C_i)$	C_i	K_j	$R(K_j, C_i)$
0	90	0.5	34	94	0.96	68	21	1.0
1	27	1.0	35	98	0.16	69	80	1.0
2	14	1.0	36	7	0.51	70	80	1.0
3	14	1.0	37	57	1.0	71	81	1.0
4	62	1.0	38	80	1.0	72	81	1.0
5	34	1.0	39	7	0.51	73	9	1.0
6	19	0.52	40	2	1.0	74	0	1.0
7	52	0.5	41	3	1.0	75	53	1.0
8	52	1.0	42	28	1.0	76	60	0.5
9	52	0.5	43	22	1.0	77	20	1.0
10	81	1.0	44	30	1.0	78	91	1.0
11	90	0.55	45	30	1.0	79	81	0.16
12	93	1.0	46	9	1.0	80	81	1.0
13	73	0.75	47	79	1.0	81	10	1.0
14	44	0.01	48	80	1.0	82	81	1.0
15	81	0.17	49	80	1.0	83	2	1.0
16	67	1.0	50	2	1.0	84	8	1.0
17	65	1.0	51	98	0.04	85	80	1.0
18	0	1.0	52	96	1.0	86	2	1.0
19	80	1.0	53	39	1.0	87	90	1.0
20	1	1.0	54	57	0.10	88	91	1.0
21	33	1.0	55	19	0.52	89	43	1.0
22	5	0.07	56	99	1.0	90	81	1.0
23	77	1.0	57	60	0.5	91	6	1.0
24	72	0.71	58	57	1.0	92	57	1.0
25	16	1.0	59	80	1.0	93	81	1.0
26	5	1.0	60	81	1.0	94	81	1.0
27	42	1.0	61	54	1.0	95	9	1.0
28	1	1.0	62	82	0.33	96	90	0.5
29	80	1.0	63	55	1.0	97	90	1.0
30	1	1.0	64	98	1.0	98	91	1.0
31	49	1.0	65	4	1.0	99	91	1.0
32	18	0.99	66	20	1.0			
33	38	1.0	67	60	1.0			

Fig. 2. The associations between the MCM and KCM

To evaluate the effectiveness of the proposed approach, it is necessary to decide the actual polarity of test messages. Since it is difficult to decide the polarity automatically, we built an online editing system to allow inspectors to label polarity score to each test message. We asked three inspectors to decide the polarity of each training and test messages. The final decisions were voted by the inspectors for each message. The result of such inspection process is used as the ground-truth polarity for the test set.

The messages in the test set were first encoded into vectors. Note that these messages may contain words that are not appeared in the vocabulary. We simply

discarded such words during the encoding process. The polarity of a test message was determined according to Eq. 6 and compared to the ground-truth polarity. We measured the accuracy over the test set and obtained the result of 85.25%.

6 Conclusions

In this work, we proposed an approach to discover the sentiments or opinions underlying social messages. First, a feature extraction process suitable for social messages is proposed. Second, a message mining process based on self-organizing map algorithm is used to discover the sentiment or opinion underlying a message. We conducted experiments using messages collecting from Twitter and obtained 85.25% of accuracy.

Acknowledgement. This work is supported by National Science Council under grant NSC 101-2221-E-390-032.

References

- [1] comSCORE: It's a social world: Top 10 need-to-knows about social networking and where it's headed (2011), http://www.comscore.com/Insights/Presentations_and_Whitepapers/2011/it_is_a_social_world_top_10_need-to-knows_about_social_networking
- [2] Nielsen: Social media report 2012: Social media comes of age (2012), <http://www.nielsen.com/us/en/newswire/2012/social-media-report-2012-social-media-comes-of-age.html>
- [3] Lipsman, A., Mudd, G., Rich, M., Bruich, S.: The power of like: How brands reach and influence fans through social media marketing (2011), http://www.comscore.com/Insights/Presentations_and_Whitepapers/2011/The_Power_of_Like_How_Brands_Reach_and_Influence_Fans_Through_Social_Media_Marketing
- [4] Krikorian, R.: New tweets per second record, and how! (2013), <https://blog.twitter.com/2013/new-tweets-per-second-record-and-how>
- [5] Kohonen, T.: Self-Organizing Maps. Springer, Berlin (2001)
- [6] Kohonen, T., Honkela, T.: Kohonen network. *Scholarpedia* 2(1), 1568 (2007)
- [7] Turney, P.D.: Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In: *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL 2002*, pp. 417–424. Association for Computational Linguistics, Stroudsburg (2002)
- [8] Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: Sentiment classification using machine learning techniques. In: *Proceedings of the ACL 2002 Conference on Empirical Methods in Natural Language Processing, EMNLP 2002*, vol. 10, pp. 79–86. Association for Computational Linguistics, Stroudsburg (2002)
- [9] Pang, B., Lee, L.: Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In: *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, ACL 2005*, pp. 115–124. Association for Computational Linguistics, Stroudsburg (2005)

- [10] Snyder, B., Barzilay, R.: Multiple aspect ranking using the Good Grief algorithm. In: Proceedings of the Joint Human Language Technology/North American Chapter of the ACL Conference (HLT-NAACL), pp. 300–307 (2007)
- [11] Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., Kappas, A.: Sentiment in short strength detection informal text. *J. Am. Soc. Inf. Sci. Technol.* 61(12), 2544–2558 (2010)
- [12] Kim, S.M., Hovy, E.: Identifying and analyzing judgment opinions. In: Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, HLT-NAACL 2006, pp. 200–207. Association for Computational Linguistics, Stroudsburg (2006)
- [13] Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2004, pp. 168–177. ACM, New York (2004)

Seek the Consent, Respect the Dissent: An Analysis of User Behaviors in Online Collaborative Community

Xiaoyue Tang¹, Hui Wang², Zhengzheng Ouyang¹, and Wei Yu³

¹ School of Mathematics and Computer Science, Wuhan Polytechnic University

² School of Information Science and Engineering, Changzhou University

³ School of Computer, Wuhan University

{sharontang,wangh,oyzz,yuwei}@whu.edu.cn

Abstract. Wikipedia is the biggest online encyclopedia whose knowledge producing process features distributed collaboration. However, its openness and tolerance of disagreements seem contrary to the objective of achieving overall consensus. To figure out what the key is to make this paradoxical editing mechanism work, we explored 202,472 historical content states for 73 articles in Wikipedia, traced the “evolution” processes of article contents, and analyzed the collaborating behaviors of the contributors from the point of content editing tendency. Finally we found that Wikipedia users tend to generalize their own expression of consensus and avoid duplicating contents from outside resource, and during their editing process, the ubiquitous initiative to approach consensus, as well as the neutral deliberation on dissent are two essential factors for collaborative communities like Wikipedia to success.

Keywords: Wikipedia, collaboration, collective intelligence.

1 Introduction

Wikipedia is a free collaborative internet encyclopedia supported by Wikimedia foundation and now with 100,000 regularly active contributors[11]. Aiming at offering a universal encyclopedia, it’s featured by its distributed collaborative editing process, in which many authors are allowed to collaborate on a single article, stored in a central database that permits easy versioning, formatting, and stylistic presentation[1]. Its 21 million articles (over 3.8 million in English) have been written collaboratively by volunteers around the world[10].

However, the proclaiming of “free encyclopedia and open to any edit” brought this online group production application tons of criticism and doubt about whether it can achieve its purpose through the distributed collaborative editing way. These negative comments are generally concerning about two unsolved problems: On the one hand, the massive emergence of false information even vandalism is continually threatening the reliability of this information democratization; On the other hand, the respect and tolerance of disagreement seems more likely to result controversy rather than consensus. “The main problem

was to ensure that the enormous, disparate community were aligned in a common cause to comprise the approval-by-consensus format of editing articles. In particular, the openness of Wikipedia brings a new salience to the challenges of consensus practice”[1]. Obviously, the multi-editor-on-single-article model inclines the editing process confronted with inevitable disagreements and conflicts, and causes obstacles in approaching coordination. Even though Wikipedia itself has its social mechanism of conflict resolution[2], the effectiveness has not been verified yet. Therefore, the dispute on the issue that collaboration in Wikipedia will finally leads to controversy or consensus has never quieted down.

Nevertheless, Wikipedia succeed after many online encyclopedia failed[12], which leaves us some basic questions: what is it presenting us eventually, a short-term content state of article or a generally approved consensus? Is it possible for an open collaborative project to achieve the expected consensus? If yes, how does the community of contributors cooperate to approach the consensus beyond all disagreements? How has the dissent inside the community been treated?

In this paper we seek to address these questions, tracing the evolution of how user-generated content changed during users’ collaboration. We analyzed 73 Wikipedia articles from two different fields: popular movies and epidemic diseases. We examined the issue of approaching consensus by comparing the article content with different approximations of the expected consensus, and we explored the general characteristic of user behaviors through the editing history. Especially, we focused on those revisions that receive negative comments from other editors. We found that the tendencies of user behaviors, that being active in approaching the mainly-approved content and being deliberate to deal with the dissent, underlines the neutral consensus generation in Wikipedia and would finally results consensus based on most editors’ approval.

2 Related Work

Since its inception in 2001, Wikipedia has drawn a great quantity of attention from researchers on various fields. A variety of issues concerning about Wikipedia and similar collaborative projects has been investigated.

Massive works have tried to explore the key factor of Wikipedia’s success. Reagle pointed out that “good faith social norms constructively facilitate Wikipedia collaboration”. Here good faith means there is a powerful norm of assuming that the person on the other side of the argument is every bit as committed as you are to getting high quality, accurate encyclopedic entries written and maintained[1]. Kittur and Kraut studied the four coordination mechanisms on managing conflicts in Wikipedia and generalized their findings to other collaborative contexts[2]. Welsler and Cosley etc. identified four key roles in Wikipedia editors and also found that informal socialization has the potential provide sufficient role related labor despite growth and change in Wikipedia[4].

There are also new methods proposed to make Wikipedia a cleaner house for collaboration. Geiger and Ribes highlighted the role of non-human actors in enabling a decentralized activity of collective intelligence[3]; to detect vandalisms

Table 1. Comparison between Movie Articles and Epidemiology Articles

Attributes	Movie Articles	Epidemiology Articles
Maximum of RPA	12395	1588
Minimum of RPA	365	175
Average of RPA	3119	598
Maximum of EPA	3864	843
Minimum of EPA	225	117
Average of EPA	1473	344
Average of RPE	1.95	1.65
Average of AR	53.20%	71.83%

automatically, Chin etc. leveraged the Active Learning and Statistical Language Models[5], Wu etc. proposed some new features based on text stability[6], Adler etc. build a content-driven reputation system to recognize the editors with bad intension[7], Koen Smets etc. used a machine learning approach[8], West etc. found that Spatio-Temporal Analysis of Revision Metadata is helpful[9].

Different from these works above, our work studied the evolution of the integral historical content states of one certain Wikipedia article, and reveals the characteristic of Wikipedians' behaviors by computing the quantified features with *NLP* method during the cooperative editing process.

3 Editing Behaviors and Approximated Consensuses

3.1 Dataset Description

Considering the differences between the editor groups of articles in different fields, we collected 196,491 historical revisions of 63 articles with 92,800 editors involved about popular movies (the first 63 movies from IMDB top 100) and 5,981 revisions of 10 articles with 3,442 editors about epidemic diseases. Regarding the possible bias of the common favor on popular culture, the editor groups of these two kinds would vary in aspects including size, expertise and editing participation. Table1 presents the comparison result between these two kinds in several basic attributes, including *Revisions per Article (RPA)*, *Editors per Article (EPA)*, *Revisions per Editor (RPE)* through one article, and *Anonymous Rate (AR)* in the editor group of an article.

3.2 Editing Behavior Representation and Consensus Approximation

We start our study on the content evolution of an article with the problem of how to estimate the editing behaviors in Wikipedia. First, we define a few critical terms. The act of making and saving changes to an article is an *edit*, and the

history of an article forms a sequence of content states called *revisions* C i.e., edits are transitions between revisions. Further, there is a special kind of edit, called *revert*: reverting an article means restoring its content to some previous revision, removing the effects of intervening edits.

Here we apply a Natural Language Processing model called *BoW* (Bag of Words) to represent the revisions. Whereas BoW does not preserve the order of words, we consider *bi-gram*, the sequence of two adjoining words, as the alternative semantic unit. Assuming the size of the bi-gram dictionary for all documents is N , if we represent the frequency of bi-gram w_j , $1 \leq j \leq N$ in a revision, R_i , as $\|w_j\|_i$, then R_i can be stated as the following N -dimension vector form:

$$\mathbf{R}_i = (\|w_1\|_i, \|w_2\|_i, \dots, \|w_N\|_i). \quad (1)$$

And we give the following vector, $\mathbf{E}_{i,i+1}$, to represent the editing move between two successional content states of an article, R_i and the following R_{i+1} :

$$\mathbf{E}_{i,i+1} = \mathbf{R}_{i+1} - \mathbf{R}_i. \quad (2)$$

Considering the possible difference between the expected true knowledge about an entry of Wikipedians and that of the outside scope, we employ three types of consensus approximation as the background knowledge for each article in our datasets: (1) The *outside resource* consensus is certified knowledge extracted from other authentic websites; (2) The *mean* article of consensus is an assemblage of the bi-grams from all historical revisions, and the frequencies of which are computed by averaging their appearing times in all the revisions; (3) The *polling* article of consensus is the aggregation comprising of all historical bi-grams, and the frequencies of them are replaced by their most frequent appearing time among all the revisions.

For articles in the dataset of popular movies, the outside resource consensus is based on the content extracted from the most award-winning movie site *IMDB* (www.imdb.com). And for articles about epidemic diseases, we approximate the consensus based on the government website *CDC* (www.cdc.gov), the representation of these outside resources is the same with that of the revisions.

Assume there are totally K revisions recorded in the editing history of an article, then the mean article of consensus C_{mean} can be stated as:

$$\mathbf{C}_{mean} = \left(\sum_{i=1}^K \|w_1\|_i / K, \sum_{i=1}^K \|w_2\|_i / K, \dots, \sum_{i=1}^K \|w_N\|_i / K \right). \quad (3)$$

To profile the polling article of consensus, we define a series of function, $D_i(x)$, $1 \leq i \leq K$, of value x , $x \in X$, to represent whether a bi-gram w appears x times in revision R_i . Here $X \subset N$ is the assemblage of all the possible frequency values.

$$D_i(x) = \begin{cases} 1 & \text{if } \|w\|_i = x \\ 0 & \text{else} \end{cases}. \quad (4)$$

Given that function $F_X(x)$ is to capture the maximum value of $x \in X$, then function $F_{1 \leq i \leq K}(\|w_j\|_i)$ of bi-gram frequency $\|w_j\|_i$ is to present the most widely appeared frequency value of w_j among all the revisions from R_1 to R_K :

$$F_{1 \leq i \leq K}(\|w_j\|_i) = \arg \max_{\|w_j\|_i} \sum_{i=1}^K D_i(\|w_j\|_i) . \tag{5}$$

Therefore, we state the polling article of consensus as following:

$$\mathbf{C}_{poll} = (F_{1 \leq i \leq K}(\|w_1\|_i), F_{1 \leq i \leq K}(\|w_2\|_i), \dots, F_{1 \leq i \leq K}(\|w_U\|_i)) . \tag{6}$$

Take article *Batman begins* for instance. There are 5,227 revisions in its editing history by Jan. 20th 2012. The bi-grams dictionary of these revisions owns a size of 47,093. Then C_{mean} and C_{poll} will be both stated as 47,093-dimension vectors. As to the elements in these vectors, consider the bi-gram *Christopher Nolan* for example. *Christopher Nolan* appears 406,728 times in all the revisions together. And among all the revisions, there're 1,739 revisions that have 86 instances of *Christopher Nolan* in their content and there's no other value of frequency polling more than 1,739 votes, thus in C_{mean} , the element for *Christopher Nolan* should be $406728/5227$, i.e., 78, and in C_{poll} , that element should be 86.

3.3 Comparison between Revisions and Consensuses

To test in what direction the collaborative editing in Wikipedia is making the contents of articles evolve, toward or away from the consensus, we need to compare the revisions to our approximations of consensus first. First we introduce a new critical term *hit*. Once a word is shown in a revision as well as in the consensus article, we say that is a *hit* of this revision to the consensus, and the word that makes a hit is called a *hitting word*. Also, we name the rest in that revision *uncertain* words. For example, assume that an original short sentence “*John likes to watch movies. Mary likes too.*” has an unanimously expected version “*John likes to watch movies. He also likes to watch football games.*” The word “*John*” makes one hit of the original revision to the consensus, the word “*likes*” makes two, and the uncertain words in the first sentence include “*Mary*” and “*too*”. Similarly we state an assemblage of hitting bi-grams in a revision R_i :

$$H_C(R_i) = \{w \mid \forall w \in C \text{ and } w \in R_i\} . \tag{7}$$

Where C stands for any type of the consensus articles. As for the rest bi-grams in R_i , the assemblage of the uncertain can be defined as:

$$U_C(R_i) = R_i - H_C(R_i) . \tag{8}$$

Which relatively suggests how much dissent (in contrast with the “hitting” content to consensus) a revision contains. Then we can state a feature called $hit_rate_C(R_i)$ of revision R_i as:

$$hit_rate_C(R_i) = \frac{|H_C(R_i)|}{|R_i|} . \tag{9}$$

In which we use the square brackets “| |” to present the size of an assemblage.

Obviously, this feature tells how much content out of all in a revision matches the consensus. Moreover, as we proposed 3 types of approximated consensus for each article, there're also 3 types of *hit_rate* characterizing matching degrees of revisions versus the consensuses, $hit_rate_{C_{out}}$, $hit_rate_{C_{mean}}$, and $hit_rate_{C_{poll}}$.

4 The Evolution of Historical Revisions

We analyzed 202,472 normal (not blank) revisions from the two datasets in section 3.1. For each article, we arranged all the historical revisions according to the posting timestamp order and successively compared them to the approximate consensuses. As the content states of an article varies along with the revisions increase, the editing history can be regarded as a process of content “*evolution*”. Then we provided a vision of the evolutionary process of the matching degree between the revisions and the consensuses for each article.

To better understand the expectation of the majority in Wikipedia collaborative groups, specially we removed those reverted revisions from the editing history and only keep the rest, then provided the comparison results between the consensus articles and the approved ones among all revisions, which are attached in the appendix at the end of this paper.

4.1 The Smoothing Effect of Reverting Behaviors

In the appendix we show the figures of the visional process that profiles how the contents of articles in the two datasets change as being continually edited by Wikipedia users. To make a whole vision of all the involved articles' content-editing histories, the articles whose contents have been edited for a lot of times are separated from those with much less historical revisions by mapping the hit rate values of frequently edited articles to higher ranges of value. Thus we divided the whole vision of the content states evolution processes for all the articles in the movie dataset into 4 sections showed in two sub-figures and divided that for articles in epidemiology dataset into 2 sections.

In Fig.4 of appendix we show the evolution process of all the revisions including those reverted ones. In all three the curves, there are a few revisions with rarely high or low hit rate values making sudden drops or rises like “*noises*”. Fig.5 illustrates the three hit_rate_C curves of not-reverted revisions only. We note that the tendencies and turning directions of the curves' shape are consistent with their parallels in Fig.4, only that the “*noises*” of sudden changes are much rarer. Since the “*noises*” are caused by massive changes in the content at one edit, this is suggesting that the edits which changes the previous content of an article by a big margin are probably reverted by their successors.

Taken together, we see that the very kind of edit, reverting, is comparatively sensitive to those revisions with great rise or drop in the *hit_rate* value when comparing to their antecedents and successors, which means, the reverting behaviors of Wikipedia editors to some extent are effective in “*smoothing*” the *hit_rate* value curves of an article's all revisions.

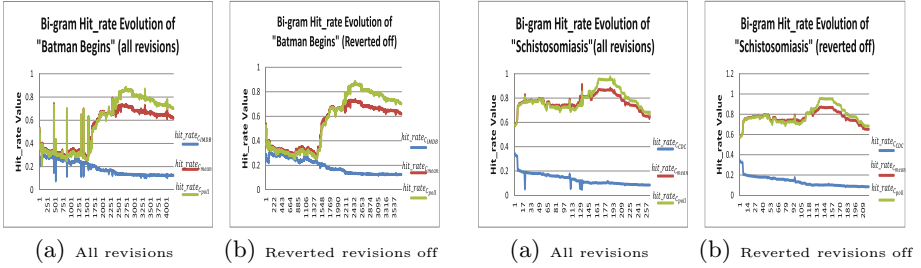


Fig. 1. hit_rate_C Evolution of *Batman Begins* **Fig. 2.** hit_rate_C Evolution of *Schistosomiasis*

4.2 The Expected Consensus of Collaboration

In Fig.1, Fig.1(a) characterizes the hit_rate evolution processes of all historical revisions of movie article *Batman Begins* while Fig.1(b) profiles those of the not-reverted revisions only. Similarly, the two charts in Fig.2 present the according analogues of the epidemiology article *Schistosomiasis*. In all the charts, the blue curves represent the changing trends of $hit_rate_{C_{out}}$ of articles, the red ones are for those of $hit_rate_{C_{mean}}$, and the green ones stand for $hit_rate_{C_{poll}}$.

Fig.1 and Fig.2 revealed a common tendency that during the evolution processes of their contents, for articles in both datasets, the content identical with the outside resource consensus are possessing a gradually decreasing part, thus the value of $hit_rate_{C_{out}}$ keeps going down slowly; in contrast to which, the contents that match the consensus merging all reversion together are increasing as the article is edited time after time, and the values of $hit_rate_{C_{mean}}$ and $hit_rate_{C_{poll}}$ are rising to the peak and then remaining in a comparatively higher range than their original after cresting over a certain high level. Whereas, although having the similar shape, the curve of $hit_rate_{C_{poll}}$ runs a higher level after reaching the peak than that of $hit_rate_{C_{mean}}$.

Revealed in both figures, there're some interesting tendencies of contributors' cooperative editing behaviors. On the one hand, although the consensus from outside resource owns widely certified authentication, it is not as favored by Wikipedians as the consensus that merges together the opinions about the article content from most editors. On the other hand, even the content coming from most editors opinions will constantly confront being questioned and modified, as the variation in curves of $hit_rate_{C_{mean}}$ and $hit_rate_{C_{poll}}$ indicated, and as the high level of $hit_rate_{C_{poll}}$ implied, only the popular bi-grams that show in most revisions are in a comparatively stable condition of being kept in the article.

Overall, the collaborative editing on an article is propelling its content towards a consensually agreed state, but also, having all editors of the same article agreed on certain content is more a dynamic process of keeping balance between approvals and disagreements.

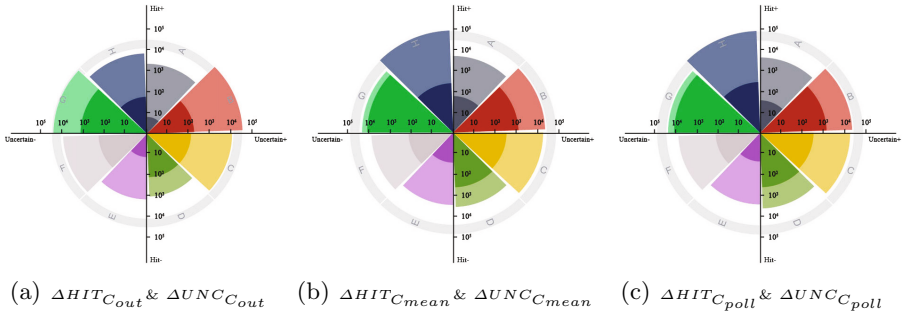


Fig. 3. Log-formed Joint Distribution of ΔHIT_C & ΔUNC_C

Table 2. Mapping Relation between Regions and Combination of ΔHIT_C and ΔUNC_C

Regions	A	B	C	D	E	F	G	H
ΔHIT_C	> 0	> 0	< 0	< 0	< 0	< 0	> 0	> 0
ΔUNC_C	> 0	> 0	> 0	> 0	< 0	< 0	< 0	< 0
$\frac{ \Delta HIT_C }{ \Delta UNC_C }$	> 1	< 1	< 1	> 1	> 1	< 1	< 1	> 1

5 Analysis and Discussion

To capture the subtle changes in the evolution of article content and reveal the tendency of editing behaviors if there's any, we define the following two metrics for an editing behavior $E_{i,i+1}$:

$$\Delta HIT_C(E_{i,i+1}) = |H_C(R_{i+1})| - |H_C(R_i)| . \tag{10}$$

$$\Delta UNC_C(E_{i,i+1}) = |U_C(R_{i+1})| - |U_C(R_i)| . \tag{11}$$

According to the 3 types of approximated consensus, we then have $\Delta HIT_{C_{out}}$, $\Delta HIT_{C_{mean}}$, and $\Delta HIT_{C_{poll}}$, as well as $\Delta UNC_{C_{out}}$, $\Delta UNC_{C_{mean}}$, and $\Delta UNC_{C_{poll}}$.

We visualised the distribution of all edits in the history of 73 articles, according to their combination values of ΔHIT_C and ΔUNC_C , and provided the demonstrations in Fig.3. In each sub-figure, the quadrants represents 4 possible combination of the ΔHIT_C value and the ΔUNC_C value for an edit, and each of them is divided into 2 regions by the line $|\Delta HIT_C| = |\Delta UNC_C|$. We showed the mapping relation between these 8 regions and the combinations of ΔHIT_C and ΔUNC_C in Table5. We use area of the vans in different regions to present the number of edits with different combinations of ΔHIT_C value and ΔUNC_C value. Specially, we deviated the reverted edit from the regular ones, divided them according to the 8 regions, and mapped the portions into the dark colored vans in the figures. Note that here all the area of vans stands for the log value of the numbers of edits in corresponding region.

Associated with the mapping relation between regions and combined values of ΔHIT_C and ΔUNC_C stated in Table5, the comparison results between Fig.3(a),

Fig.3(b) and Fig.3(c) reveal some contributing factors of the tendencies in collaborative behaviors:

Factor 1. The light color vans have more balanced areas in Fig.3(a), the large ones appear almost evenly in region B, C, F, and G, while in Fig.3(b) and Fig.3(c), apparently the largest light color vans are shown in region H. That is to say, the largest proportion of edits contribute the hits to C_{mean} or C_{poll} into the article content, and also reduce the uncertain bi-grams to C_{mean} or C_{poll} , which causes the tendency of the article content to evolve into the state of a mostly approved consensus rather than an authentic article from outside.

Factor 2. The smallest light color vans in Fig.3(b) and Fig.3(c) are both located in regions F, while that in Fig.3(a) is in region D. This implies that only very few edits have the intention to reduce both hitting bi-grams and uncertain ones to C_{mean} or C_{poll} , especially the hits, which gives the explanation in another aspect of the tendency that the article content are changed into the consensus inside the editing group as long as the article length increases gradually.

Factor 3. The largest dark color vans are shown in region B and G in both Fig.3(b) and Fig.3(c), while the dark color vans in Fig.3(a) have much vaguer partitions in the area value. This is suggesting that most of the reverted edits are more concerned in modifying (either add or delete) the uncertain bi-grams to C_{mean} or C_{poll} than the hitting bi-grams, which however is not true for the uncertain bi-grams to C_{out} . In another word, the edits that cares not-generally-approved content more than the generally-approved content are of the best possibility to be reverted, even if their intention was to delete more uncertain content.

As we mentioned in section 3.3, the uncertain content relatively represents the dissent in contrast with the hits, therefore, besides the smoothing effect of reverting behaviors, this tendency also reflects a neutral deliberative attitude of the editors in dealing with contents that is different from the consensus.

Taken together, we find that in wikipedia, there're two valuable prevalent attempts lying in the regular edits: one is the constant effort of concentrating the article content into the consensus approved by all the editors (or at least most), the other is the reasonable deliberation on the dissent content.

6 Conclusion

Our research began by suggesting that the common expectation of all editors to achieve consensus is a key to the success of an online collaborating community like Wikipedia. Through an exploratory analysis on historical revisions of Wikipedia articles, we proposed three possible methods of approximating the consensus, defined some linguistic features to measure the matching degree of revisions versus the approximated consensus, and provided a preliminary illustration of the content evolution processes for different articles. In addition, we discussed the cause of those tendencies from the aspect of user behaviors and explained the changing patterns of an article's content.

There is much room left for improvement in future research. First, collaborating pattern is another key that affects the quality of a group producing project, and more factors should be taken into account to find those patterns. Second, more features of revisions and editing behaviors need to be explored and applied to the practical applications like vandalism detection or editing robot's function improving. Wikipedia is a good object of study for researches on online community, especially regarding the open data it provides to public and its amazing mechanisms of keeping such a huge collaborating group in order.

In conclusion, we found that achieving consensus is the key to keep collaborative projects working, specially in Wikipedia community, the editors prefer generalizing the expression of consensus of their own to absorbing authentic content from outside resource. Moreover, there're two essential factors for the success of collaboration to conquer all disagreements, one is the active efforts to approach consensus, and the other is the neutral deliberation on dissent.

Acknowledgments. We would like to thank Qiaozhu Mei, Amanda Nemo, Tao Sun, and countless Wikipedia contributors.

References

1. Reagle, J.M.: Good Faith Collaboration – The Culture of Wikipedia (Web edition). The MIT Press, Cambridge (2011)
2. Kittur, A., Kraut, R.E.: Beyond Wikipedia: Coordination and Conflict in Online Production Groups. In: CSCW 2010, Savannah, Georgia, USA (2010)
3. Geiger, R.S., Ribes, D.: The Work of Sustaining Order in Wikipedia: The Banning of a Vandal. In: CSCW 2010, Savannah, Georgia, USA (2010)
4. Welser, H.T., Lin, A., Cosley, D., Dokshin, F., Smith, M., Kossinets, G., Gay, G.: Finding Social Roles in Wikipedia. In: iConference 2011, Seattle, WA, USA (2011)
5. Chin, S., Street, W.N., Srinivasan, P., Eichmann, D.: Detecting Wikipedia Vandalism with Active Learning and Statistical Language Models. In: WICOW 2010, Raleigh, North Carolina, USA (2010)
6. Wu, Q., Irani, D., Pu, C., Ramaswamy, L.: Elusive Vandalism Detection in Wikipedia: A Text Stability-based Approach. In: CIKM 2010, Toronto, Ontario, Canada (2010)
7. Adler, B.T., Alfaro, L.: A ContentDriven Reputation System for the Wikipedia. In: WWW 2007, Banff, Alberta, Canada (2007)
8. Smets, K., Goethals, B., Verdonk, B.: Automatic Vandalism Detection in Wikipedia: Towards a Machine Learning Approach. In: 2008 Association for the Advancement of Artificial Intelligence (2008)
9. West, A.G., Kannan, S., Lee, I.: Detecting Wikipedia Vandalism via Spatio-Temporal Analysis of Revision Metadata. In: EUROSEC 2010, Paris, France (2010)
10. Wikipedia in Wikipedia extracted (March 6, 2012), <http://en.wikipedia.org/wiki/Wikipedia>
11. Technology can topple tyrants: Jimmy Wales an eternal optimist. Sydney Morning Herald (November 7, 2011)
12. Giles, J.: Internet encyclopedias go head to head. Nature 438, 900–901 (2005)

Appendix

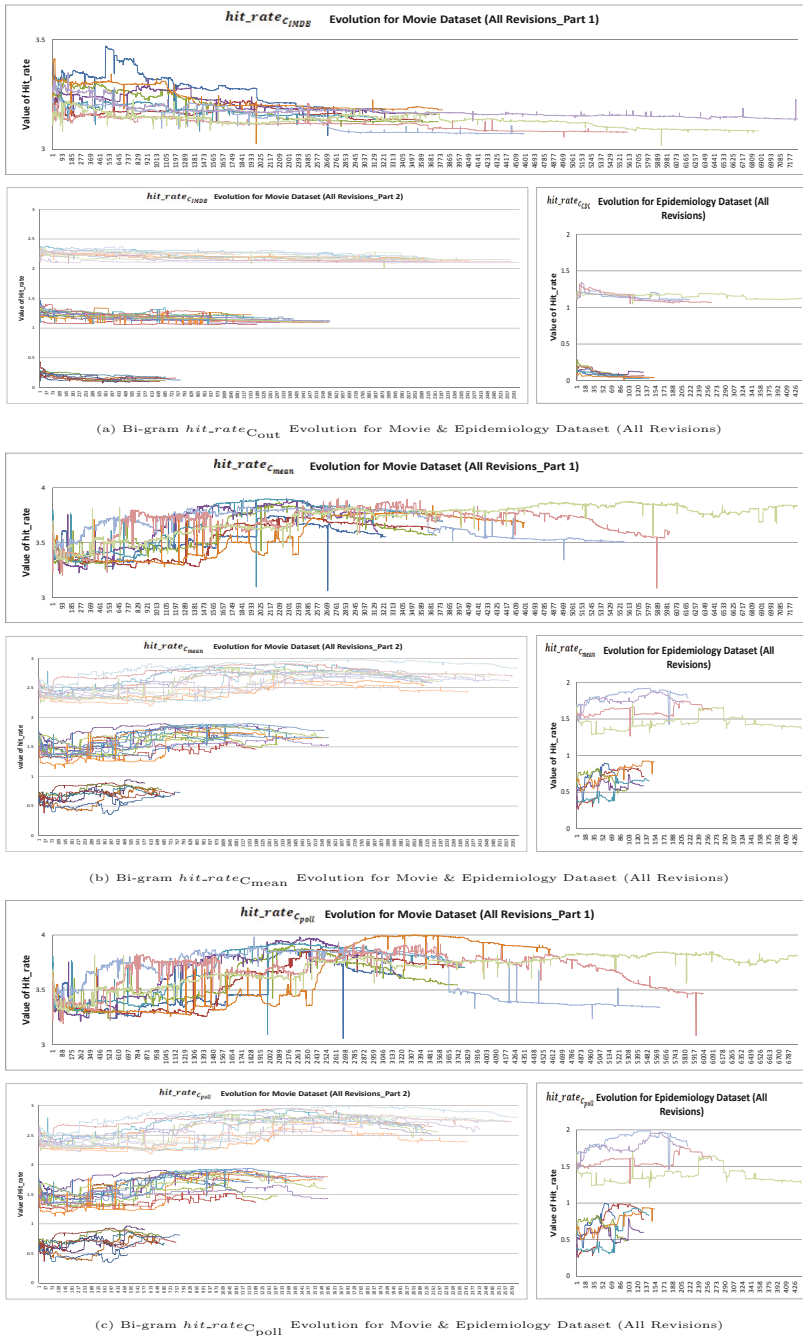
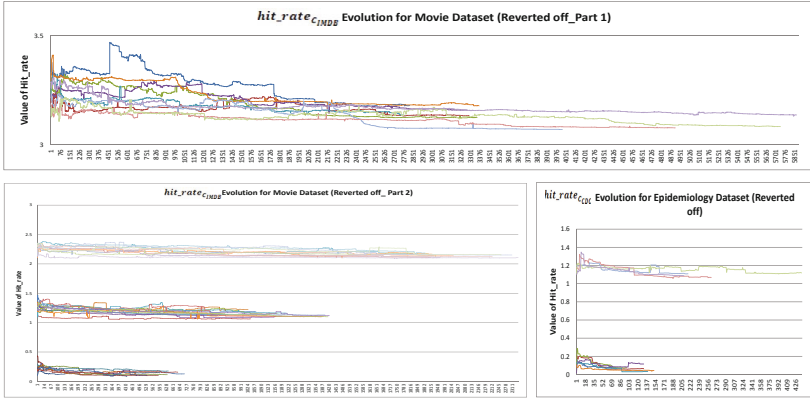
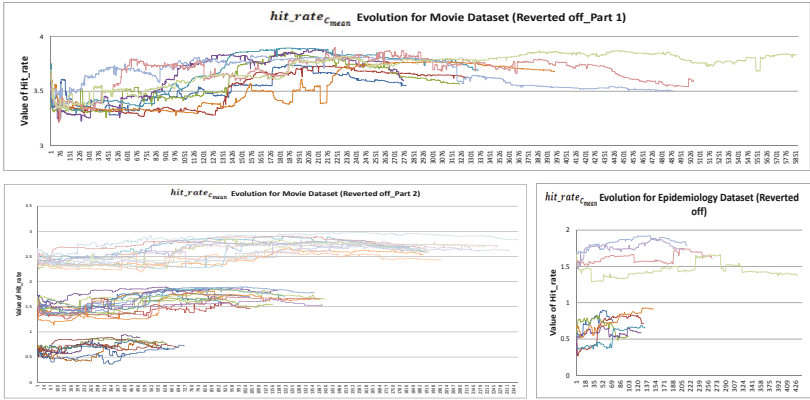


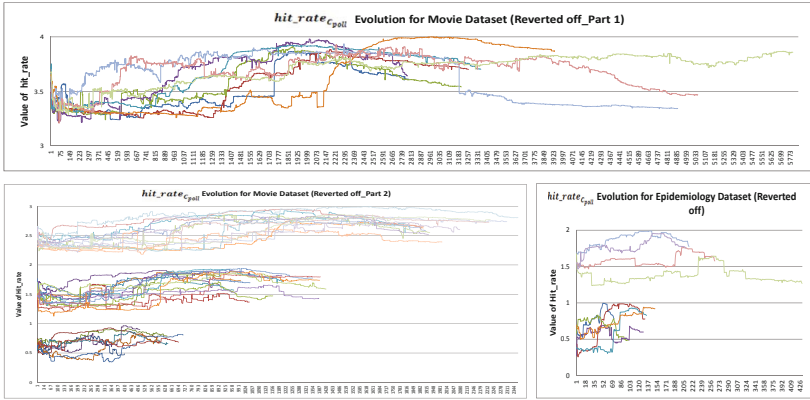
Fig. 4. Bi-gram hit_rate_C Evolution for Movie & Epidemiology Dataset (All Revisions)



(a) Bi-gram $hit_rate_{C_{OUT}}$ Evolution for Movie & Epidemiology Dataset (Reverted off)



(b) Bi-gram $hit_rate_{C_{mean}}$ Evolution for Movie & Epidemiology Dataset (Reverted off)



(c) Bi-gram $hit_rate_{C_{poll}}$ Evolution for Movie & Epidemiology Dataset (Reverted off)

Fig. 5. Bi-gram hit_rate_C Evolution for Movie & Epidemiology Dataset (Reverted off)

Study on Parallax Scrolling Web Page Conversion Module

Song-Nian Wang* and Fong-Ming Shyu

Department of Multimedia Design, National Taichung University of Science and Technology
phenombox@gmail.com, fms@nutc.edu.tw

Abstract. Parallax scrolling is one of the popular web page design effects in recent years. It enhances web page to convey the message more clearly than static pictures and text only. Many parallax scrolling webpages are designed by reconstructing a new production with open source designing tools.

This study attempts to convert non-parallax web page into parallax through the parallax scrolling conversion module that developed by this study. This module can easily convert existing web page into parallax scrolling web page. The first step of this study is to connect the *skrollr* parallax scrolling link tag with jQuery libraries and CSS into html tag of original web pages. Next step, web designers just have to input the affected range (such as a HTML or ID name, and so on) by parallax scrolling module developed by this study to generate *skrollr* parallax scrolling code. In the final step, paste the generated code back to the original web code and then the generating processes of parallax scrolling web page are completed.

Keywords: web page conversion, parallax scrolling, web effect, jQuery.

1 Introduction

Early Web design combined with pure text form and static pictures as the main content. After the Web technology matures began to appear animation, video, audio and other interactive media effects. Web browser use Flash, Shockwave and QuickTime plug-ins and other support that can present quality dynamic vision effects, until today still favored by many commercial web pages. [1]

In recent years affected by HTML5 technology with JavaScript and CSS3 are able to use the plug without display rich vision effects and more applications to multimedia content in web design such as banner animation, parallax scrolling, and video. Parallax scrolling is a common on HTML5 website, and using web layers and objects of different rate of movement to generate staggered visual effects. Figure 1 shows typical parallax scrolling web page that the whole picture is scrolling from top to the bottom. The speeds of each planet are moving differently. The planets of distance are approaching due to moving upward faster or the planets of distance are farther due to moving upward slowly.

* Corresponding author.

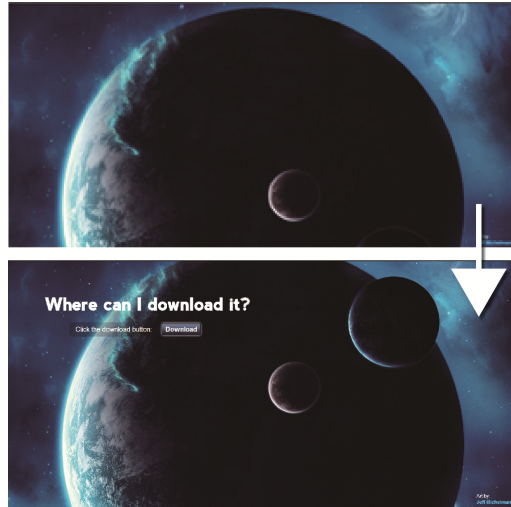


Fig. 1. Parallax Scrolling Web Page Schematic (Source: <http://www.jarallax.com/demo/>)

Many parallax scrolling web pages are designed by re-constructing a new production with open source design tools. We have developed a conversion model to semi-automatically construct parallax scrolling pages through manipulating HTML and CSS only. The kernel module is parallax scrolling open source *skrollr* library. The first step must enter code into the HTML tag connection *skrollr* and link style files in the original web page. Secondly, we need to understand HTML tags, attributes of the original web page, and identify the area affected to apply parallax scrolling effect. Finally, the designers just have to modify the parameter values to arrange the web page such as resizing, transparent, rotation angle. With scrolling object position display at different speeds on web page that is the primary features of parallax scrolling.

2 Related Work

2.1 HTML

HyperText Markup Language is the main markup language for creating webpages that develops open standards by the World Wide Web Consortium (W3C). HTML4.01 is the W3C Recommendation version and HTML5 is the next major revision version of the HTML standard working group as an updated revision to the candidate Recommendation. However, there have been many browsers support HTML5, such as Firefox, Chrome, Safari, Opera and Internet Explorer 9 later versions. HTML5 syntax followed HTML4.01 addition and improvement of the tag and the Application Programming Interface (API) to create a web application and handling the document object model (DOM) [2]. The following are the main technologies of HTML5:

- (1) Canvas Tags: Canvas is used to draw graphics via JavaScript that is able to render complex scenes and timely computing animation.

- (2) Positioning: Support mobile positioning on the device, the browser which can obtain the user's location.
- (3) Multimedia tags: It is able to define videos and audio by video and audio tag.
- (4) Enhance definition of tag: HTML5 for the HTML4.01 tag do further addition new tag such as header, footer, section and article.
- (5) Offline storage: Program information write to the client browser cache even if leaving the main page, turning back to the browser the program still can write and access to the browser's cache next time.

2.2 CSS

Cascading Style Sheets (CSS) is a standard developed by W3C to define the appearance of web documents. It can be used to conveniently set the text font and color, the location of the picture, the size of the form, and graphic layout. CSS styles and pages can be placed in different files separately and make web design become so simplified and modular.

CSS3 is the latest version of the standard early as 1999 had already begun to develop until June 7 in 2011 became a W3C recommended specifications. CSS3 adds some new features, such as rounded corners, shadows and transparency, beautify richer web pages, but the drawback is too old browsers to render.

2.3 jQuery

jQuery is an open source by the John Resig on the 2006 release of the BarCamp NYC. JavaScript is the most widely used library that the purpose of working a lot of things with small amount of code. It makes many things easy to operate such as creating animation, event handling, and select DOM documents. Making each function return value is the element itself; therefore, it can be reached through a series of continuous processing function. jQuery has some other interface plug-in module which are jQuery UI, jQuery Tools, jQuery Mobile and so on.

2.4 Parallax Scrolling Web

Parallax scrolling web page is to achieve multi-layered background via mouse wheel or drag through the browser scroll bar to move at different speeds, the formation of three-dimensional motion visual effects. Parallax effect in web design has been widely used in recent years. The elements of the page in each layer have different moving speed when the user rolls the mouse wheel. On the one hand, the faster moving speed of elements is in upper layer. On the other hand, the lower moving speed of elements is in lower layer. Visitor's visualization is different from the past of the visitor's browsing experience. Parallax scrolling is usually based on a single web page, the higher value of web pages content is plentiful, and visualized effects are more obviously than before. Most parallax scrolling web design are applying JavaScript as open source such as *Stellar.JS*, *SuperScrollerama*, *Jarallax*[4] and *skrollr*, thus web designing will be produced in different ways.

Parallax pages is the typical way to present by scrolling or moving the mouse, the speed of objectives are moving differently, showing three-dimensional foreground and background, but there are other manifestations. The following are different ways to present the case parallax scrolling web pages. Figure 2 shows parallax scrolling via mouse wheel to scroll and showing a horizontal scroll animation, and breaking the traditional web browsing thinking. Figure 3 shows parallax scrolling web pages hiding the scroll through the mouse wheel button or touch the sliding direction. While users browse web content to multiple steps one by one, web page shows the product from different angles with add text to enhance web browsing effects.

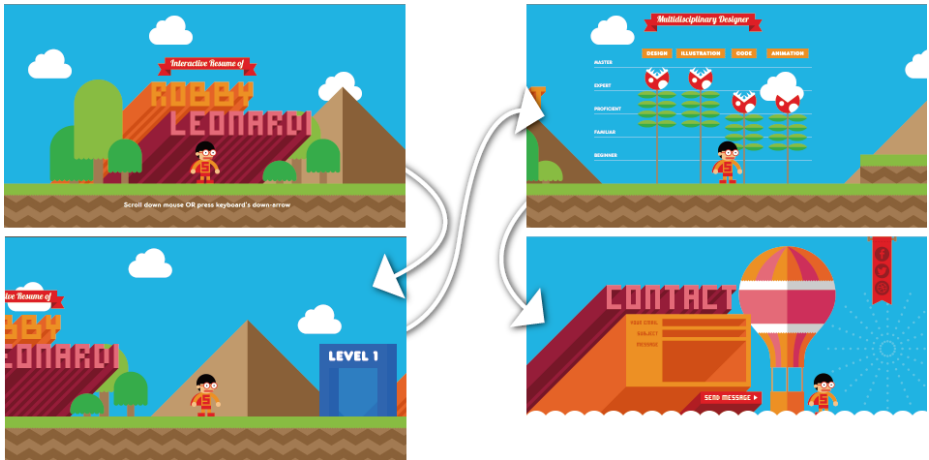


Fig. 2. horizontal scroll parallax scrolling web pages (Source: <http://www.rleonardi.com/interactive-resume>)

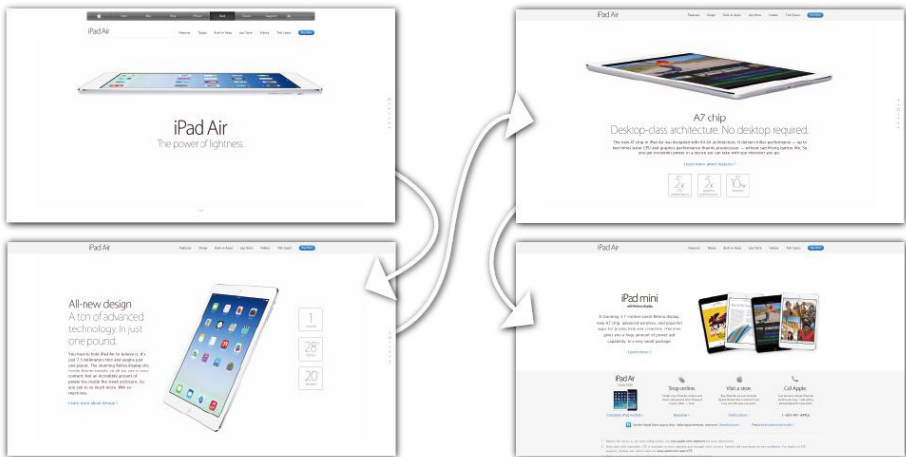


Fig. 3. Product parallax scrolling web pages (Source: <http://www.apple.com/tw/ipad-air/>)

3 Customizable Components Solutions

The methods of this study are conveniently and quickly converting non-parallax scrolling effect of the web pages into parallax scrolling web page through modifications simply. HTML tags are written in the original web page to import an external JavaScript and adjust parallax scrolling attribute parameters to custom JavaScript file. The attribution will affect the CSS of original page while scrolling web page. The CSS web page is going to convert as scroll position then you can create a variety of different parallax scrolling effect, hereinafter referred to as the parallax conversion module. While the corresponding tag or attribute name. You can quickly modify the rendering of parallax scrolling web pages by using developed module of this study. Furthermore, One Tea shop's web page as the sample in this study is to demonstrate the processes of conversion.

3.1 Original Page Framework

Determined of modification web pages, we need to understand what the page using tag, id, or class attributes such as name firstly. The purpose is to generate corresponding code of custom modules. This study will begin with id affected by parallax scrolling on the page name respectively. The outer layer is wrapper, and the left side is linkbar and the right side is page and aboutIMG. Figure 4 shows effects of parallax scrolling by the id attribute.

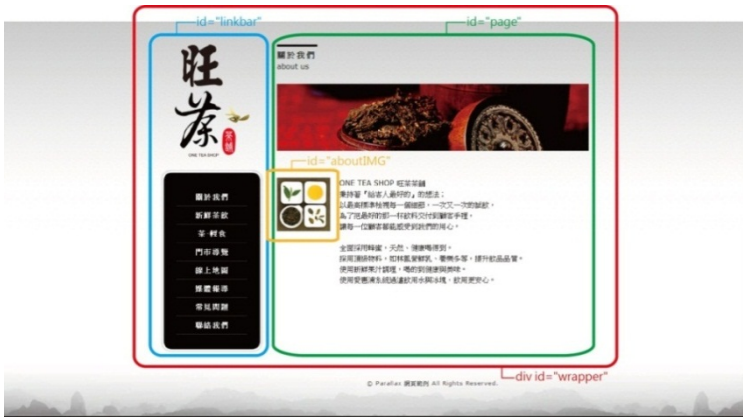


Fig. 4. The example original page attribution naming

3.2 Conversion Process

As Figure 5 system process, the original page using jQuery, *skrollr* libraries, and needed CSS file. Regards to produce code via parallax scrolling module, it needs to match with converting section in order to copy in the name of the original page. It modifies the parameters of the scrolling position and web object's attribute. For example, about.html page has an id attribute "page", and in the customized code using jQuery page corresponding id name tag. Modifying web pages into parallax scrolling

parameters needed, such as scrolls to 0, the object on the left screen, scrolling to 1000, the object moves to the right screen. In above parameters of scrolling presentation and div tags can be modified at any time.

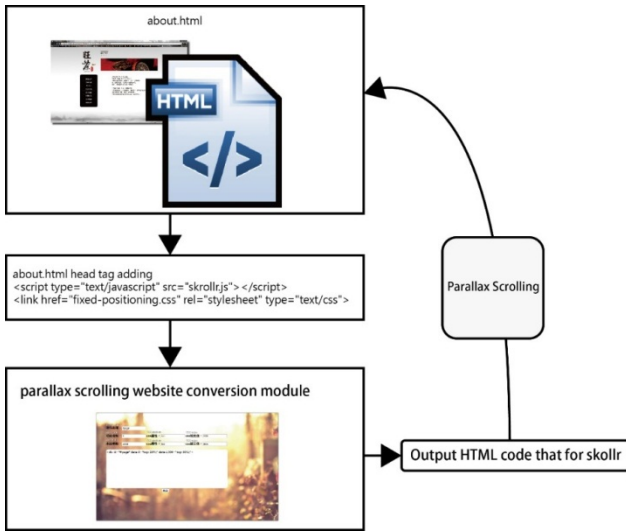


Fig. 5. Conversion Process

4 Conversion Module Implements

This study implements web page conversion prototype using Adobe Dreamweaver CS6, jQuery, and the *skollr* library. jQuery is a JavaScript open-source library, it can

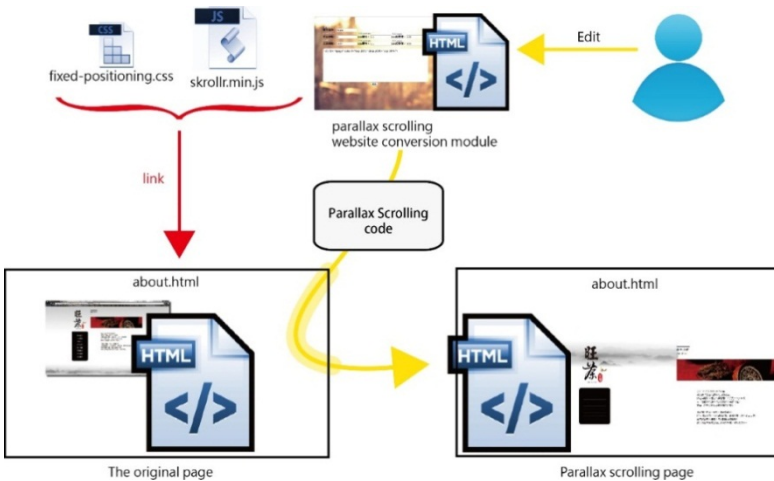


Fig. 6. System Architecture

simplify the coding. Using jQuery selector can easy to transfer HTML tags. There are many kinds of parallax scrolling web pages libraries. In this study we selected *skrollr* as an example. The following describe the results of this study demonstration and implementation prototype results. In this study, operations of the system architecture are shown in Figure 6. Original page requires link with fixed-positioning CSS file and *skrollr* JavaScript file.

When *skrollr* JavaScript and fixed-positioning style files connected to the original page, we can edit the parallax conversion module. In Figure 7, the parallax conversion modules are transmitted using jQuery tags and attributes from the original page. The tag of the corresponding original page has the same id name, adds attributes and parameters in a div tag. The scroll position data-0 indicates at the top of the reel and data-1000 sets reel at the bottom. Numerical data is not limited to the bottom, and the distance of scroll will be longer due to great value. The top or right are CSS attributes. Besides, controlling the position of the object can also control the length, width and the transparency of scrolling changes that produced a parallax scrolling effect. Therefore, in the same web page can develop different parallax scrolling effect. Developers only need to modify the attribute and parameters parallax conversion module. They can develop a variety of parallax scrolling web pages module for further usages.

The screenshot displays a web-based editor for the parallax conversion module. It features several input fields and a text area:

- HTML id:** 物件名稱: #page
- scroll initial:** 初始捲軸: 0
- scroll end:** 末端捲軸: 1000
- CSS attribute:** css屬性: top
- CSS value:** css設定值: 20%
- CSS attribute:** css屬性: top
- CSS value:** css設定值: 80%

The text area contains the following HTML code:

```
<div id="#page" data-0="top: 20%;" data-1000="top: 80%;">
```

At the bottom right, there is a button labeled "輸出" (Output).

Fig. 7. Parallax conversion module editor explanation

In this study, One Tea Shop web page as an example, the conversion processes of the original web page into parallax scrolling web page are shown in Figure 8. When the scroll position on the top, the data-0 is show in the upper left. Meanwhile the screen of object has not appeared yet. When the scrolling is going to down, you will find the object to move center of screen. The scrolling is going to the bottom which is the data-1000 and parallax scrolling animation ends. The data parameter values will affect the length of scrolling drag. Completion of parallax scrolling pages uploaded to the server, URL:

<http://163.17.137.161/ECC2014/about.html>.



Fig. 8. Schematic view parallax scrolling web animation after the conversion

5 Conclusion and Future Work

In recent years, parallax scrolling techniques are more favored by the developers of business web page. Various presentation of parallax scrolling web pages can be seen on the network. With more new design, this study uses a relatively simple way and develops a module to add parallax scrolling effect on original web pages. This study attempts to use application with jQuery and *skrollr* library to develop parallax scrolling conversion modules. Using this module, web developer can easily add parallax scrolling effect to the original web page. According to the framework of this study, web developer is able to design multiple of parallax scrolling web pages module and easy to render parallax scrolling effect.

The parallax scrolling web conversion module has been developed in preliminary stage. Original web pages developers still need to update the internal HTML structure manually. This inconvenience problem can be solved by using plug-in or off-line web update in the next version of module. The module of this study expected to be combining with mobile web page due to the rapid growth of mobile devices.

References

1. Anttonen, M., Salminen, A., Mikkonen, T., Taivalsaari, A.: Transforming the Web into a Real Application Platform: New Technologies, Emerging Trends and Missing Pieces. In: SAC 2011, March 21-25, TaiChung, Taiwan (2011)
2. Chen, W.A.: From HTML5/CSS3/Javascript to jQuery/PhoneGap Android program design. Flag Corp. (2012) ISBN 978-986-312-038-4
3. Wei, C., Lee, H., Molnar, L., Herold, M., Ramnath, R., Ramanathan, J.: Assisted Human-in-the-Loop Adaptation of Web Pages for Mobile Devices. In: IEEE 37th Annual Computer Software and Applications Conference, pp. 118–123 (2013)

4. Jarallax open source advanced javascript animation library (November 25, 2013),
<http://www.jarallax.com/demo/>
5. Interactive Resume of robby leonardi (January 17, 2014),
<http://www.rleonardi.com/interactive-resume>
6. Apple Inc. (January 20, 2014) <http://www.apple.com/tw/ipad-air/>
7. W3Schools Online Web Tutorials (December 25, 2013),
<http://www.w3schools.com/>

A Graph Theory-Based Evaluation of Strategy Set in Robot Soccer

Jie Wu¹, Václav Snášel², and Guangzhao Cui¹

¹ Department of Electrical Engineering,
School of Electric and Information Engineering,
Zhengzhou University of Light Industry, Zhengzhou 450002, P.R. China
defermat2008@hotmail.com, cgzh@zzuli.edu.cn

² Department of Computer Science,
Faculty of Electrical Engineering and Computer Science,
VŠB – Technical University of Ostrava, Ostrava 70032, Czech Republic
vaclav.snasel@vsb.cz

Abstract. Strategy evaluation in robot soccer is a very important issue in the field of multi-robot coordinated control system. In our work, the implication of strategy and strategy set in robot soccer game are described firstly, it helps to explain the morphology of strategy set and make clear the four types of strategy subset. By transferring strategy set to a directed graph, we present a directed graph-based approach to evaluate the strategy set in robot soccer game. In this idea, better strategy set would achieve higher probability of goal score, then the probability of goal score is the benchmark of strategy set evaluation. According to the directed graph of strategy set, a group of linear equations can be constructed to compute the probability of goal score. In order to testify our method, two strategy sets are evaluated, and twenty simulation games are played to compare the performance of two strategy sets, the results of twenty games validate our approach.

Keywords: graph theory, strategy set, robot soccer.

1 Introduction

Robot soccer is a classic integration of artificial intelligence (AI) and robotics, where AI has become one of the most attractive fields, and robotics represents the advanced level of science and technology currently. As we known, soccer is a team sport. In human soccer, the collaboration is very important. Similarly, in robot soccer, effective collaboration ensures victory. This collaboration requires not only cooperation between robots, but also collaboration with high efficiency.

In the robot soccer game, robots stay in a dynamic environment all the time. The teammate robots need to make decisions and take actions according to the changing situation, to gain ascendancy and win the game by means of teamwork. In this process, the right decisions and appropriate actions give teammates advantage over opponent. Accordingly, we can divide the robot soccer game into

strategy level and *tactic* level. On the strategy level, the robots need to make the right decisions based on game situation. On the tactic level, the robots need to take actions to implement the strategies. In a sense, the effective collaboration in robot soccer depends on the organization of teamwork, while the strategy represents this kind of teamwork organization.

A strategy in human soccer game is a plan to accomplish the game goals [4]. Every player will follow its strategy to get the maximum possible advantage for himself. A plan consists of choices, which includes condition and decision. The condition means current situation, the decision represents the actions taken by the player as a response to the current situation for getting the maximum possible advantage for himself. Similarly, a strategy in robot soccer is a plan, or called decision-making, which could be expressed by “if-then” statement including condition part and decision part.

There are many techniques have been applied to decision-making of robot soccer game, including Case-based Reasoning [22–24], Learning from Observation [8,12], Reinforcement Learning [11,21], Pattern Recognition [9,13,16], Fuzzy Theory [14,25], Neural Network [10], Evolutionary Algorithm [19,20], etc. Lots of decision-making mechanism in robot soccer system are based on experience, which can neither evaluate the decision-making itself nor answer the question that how good are the strategies. For instance, to solve a new problem in a changing environment, if a new decision is made based on a “weak” decision, then the result is likely to be “weak”. Thus, it is very important to set up an approach to evaluate the strategy in robot soccer.

The paper is organized as follows. The strategy and strategy set in robot soccer is described in Section 2 and Section 3 respectively. Section 4 presents a novel approach to evaluate the strategies in robot soccer, and twenty games are played to testify our approach. Finally, Section 5 draws the conclusions.

2 Description of Strategy in Robot Soccer

The strategy is a plan which could be expressed by an “if-then” conditional statement, where the condition part describes the current situation, and the decision part declares the response actions. The current situation is usually represented by a rough subregion position. We preserved all these essential elements in our strategic description.

Browning *et al.* [1–3] presented a hierarchical architecture, called STP (i.e. acronym of skills, tactics, and plays). In this hierarchical architecture, a play could be conceived as a strategy-level plan, the tactics and skills are implementation of the play, and tactics for abstract while skills for specific.

Lucchesi [15] and Dylla *et al.* [5–7] define the strategy as a tuple $str = (RD, CBP)$, where RD is a set of *role descriptions* that could be conceived as a set of tactics, CBP is the set of *complex behavior patterns* associated with the strategy that could be considered as the corresponding actions.

Ros *et al.* [23] applied case-based reasoning (CBR) techniques into robot soccer. The case definition is composed of three parts: the problem description P , the

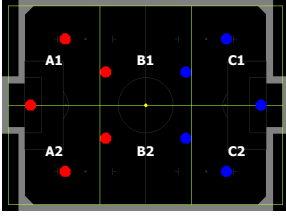


Fig. 1. An Example Case

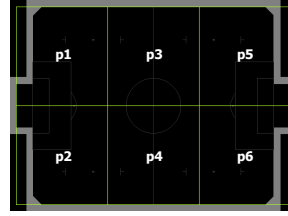


Fig. 2. A Strategy Grid

solution description A , and the case scope representation K . Then they formally define a case as a 3-tuple $case = (P, A, K)$. Therefore in their work, the problem description P is the current condition of strategy, the solution description A is the corresponding tactics.

Nakashima *et al.* [17–19] proposed an evolutionary method to improve team strategies for soccer agents. They defined ten actions for the consequent action C_j , the value of C_j corresponds to the index number of ten actions. They use an integer string of length 960 to represent a rule set of action rules. Therefore, their aim is to evolve the integer strings then obtain team strategies with high performance. The performance is measured by the scores of soccer games.

All the methods mentioned above have much in common. First, a strategy is a plan. Secondly, all the strategies could be expressed by “if-then” statement. Thirdly, in the condition part of the statement, the strategy is associated with a subregion of the playground. Fourthly, in the decision part of the statement, the strategy declares a sequence of actions which could be conceived as tactics. All these common features helps to describe our robot soccer strategy in our work.

According to these common features of robot soccer strategy, we express the game situation only by strategy grid position. Therefore, the strategy can be expressed easily as (M, O, B, D) , where M is the teammates’ positions of *mine*; O , *opponents’* positions; B , *ball* position; and D , *my* teammates’ *destination* grids.

Fig. 1 shows a strategy stored in a log file. It means “**If** (M_1, M_2, M_3, M_4) is close to $(A1, A2, B1, B2)$, **and if** (O_1, O_2, O_3, O_4) is close to $(B1, B2, C1, C2)$, **and if** B is close to $(B2)$, **then** (M_1, M_2, M_3, M_4) go to $(A1, B2, C1, C2)$ ”. In the strategy, the first part (M, O, B) is *condition attributes*, the latter part D is *decision attribute*.

In the field of strategy, there are two types of features, i.e. controllable features and non-controllable features [23]. Teammates’ positions are controllable features, while the ball’s and opponents’ positions are non-controllable features. Here we ignore the ball position when we analyze the strategy set, because the ball position must be in accordance with one of teammate or opponent robots’ position. By the control tag of the ball, the strategies can be divided into two sets which correspond to attack and defence respectively. It is easy to switch the team’s state between attack and defence in this way. Consequently, according to the strategy grid in Fig. 2, the strategy in Fig. 1 can be simplified as (111100100111) that means “**if** M is close to $(p1, p2, p3, p4)$ **then** M goes to $(p1, p4, p5, p6)$ ”. Here M means (M_1, M_2, M_3, M_4) .

Table 1. A Strategy Set

U	c_1	c_2	c_3	c_4	c_5	c_6	d_1	d_2	d_3	d_4	d_5	d_6
x_1	0	0	1	1	1	1	0	1	0	1	1	1
x_2	0	1	0	1	1	1	0	1	1	1	1	0
x_3	0	1	1	0	1	1	0	0	1	1	1	1
x_4	0	1	1	1	0	1	1	1	1	0	0	1
x_5	0	1	1	1	1	0	1	0	0	1	1	1
x_6	1	0	0	1	1	1	1	0	1	0	1	1
x_7	1	0	1	0	1	1	0	1	1	0	1	1
x_8	1	0	1	1	0	1	1	0	1	1	0	1
x_9	1	0	1	1	1	0	1	1	0	0	1	1
x_{10}	1	1	0	0	1	1	1	0	1	1	1	0
x_{11}	0	0	1	1	1	1	0	1	0	1	1	1
x_{12}	1	1	1	0	1	0	0	0	1	1	1	1
x_{13}	1	0	0	1	1	1	1	1	0	1	0	1

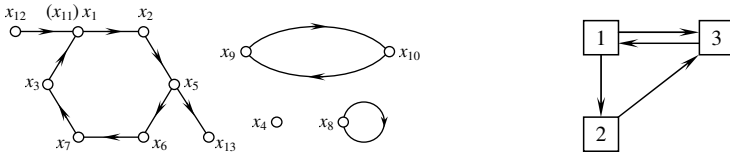


Fig. 3. Morphology of Strategy Set in Table 1 **Fig. 4.** A digraph of a strategy set

3 Strategy Set in Robot Soccer

A number of strategies constitute strategy set. Strategy set is the kernel of teamwork in robot soccer game. A strategy set could be represented as a directed graph, or for short a digraph, in which a vertex represents a strategy (or a rule), the direction of edge indicates that executing the former strategy would lead to executing the latter strategy. Table 1 shows a strategy set containing 13 rules. The corresponding directed graph is depicted in Fig. 3.

According to the directed graph, we can find that there are three types of strategies in strategy set. In Fig. 3, x_4 is an *isolated strategy*, which means there is no other strategy connects up x_4 , and x_4 does not connect up any other strategy; $x_{12} - x_1 - x_2 - x_5 - x_{13}$ constitute *chain strategies*; $x_1 - x_2 - x_5 - x_6 - x_7 - x_3 - x_1$ constitute *loop strategies*.

About the loop strategies, there are three types again. The first one is *self-loop strategy*, such as x_8 ; the second type is two-component loop strategies which are constituted by two strategies, such as x_9 and x_{10} ; the third one is formed by more than two strategies, called multi-component loop strategies, such as $x_1 - x_2 - x_5 - x_6 - x_7 - x_3 - x_1$. If we cast off any one strategy in the third type, we can obtain a chain strategies.

In the directed graph of Fig. 3, x_{11} is same to x_1 , they are superfluous rules; x_{13} and x_6 have the same condition attribute but different decision attribute, therefore they are inconsistent rules.

There are 4 types of subsets for robot soccer strategy set. Two of them are for teammates and two for opponents, two for attack and two for defence. The corresponding sorting conditions are listed as follows.

1. Rules Subset TA (Teammates–Attack) : (ball is close to teammate robots) || ((ball is NOT close to teammate robots) && (ball is NOT close to opponent robots)).
2. Rules Subset TD (Teammates–Defence) : (ball is close to opponent robots) && (ball is NOT close to teammate robots).
3. Rules Subset OA (Opponents–Attack) : (ball is close to opponent robots) || ((ball is NOT close to teammate robots) && (ball is NOT close to opponent robots)).
4. Rules Subset OD (Opponents–Defence) : (ball is close to teammate robots) && (ball is NOT close to opponent robots).

A directed graph is a kind of representation of strategy set, in which the node can be considered as an *event*. Event i means the robots stand at the position (or state) corresponding to node i . Each arrow pointing to event j from event i represents a rule, r_{ij} , that the decision attribute would be position j under the condition of position i . The probability of event i , denoted by $P(i)$, means the probability of robots' standing at the position i . The probability $P(j|i)$ denotes the probability of rule r_{ij} being selected. For example, a strategy set could be represented as the directed graph shown in Fig. 4. According to Fig. 4, we have following equations.

$$P(1) = P(1|3)P(3) \tag{1}$$

$$P(2) = P(2|1)P(1) \tag{2}$$

$$P(3) = P(3|1)P(1) + P(3|2)P(2) \tag{3}$$

$$P(1) + P(2) + P(3) = 1 \tag{4}$$

$$P(1|3) = 1 \tag{5}$$

$$P(3|2) = 1 \tag{6}$$

$$P(2|1) + P(3|1) = 1 \tag{7}$$

These linear equations can be written as

$$\mathbf{Ax} = \mathbf{x}, \tag{8}$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & P(1|3) \\ P(2|1) & 0 & 0 \\ P(3|1) & P(3|2) & 0 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} P(1) \\ P(2) \\ P(3) \end{bmatrix}.$$

According to Eq. (8), we can find that if the node i in directed graph means the state of goal, then we can compute the probability of goal.

4 Evaluation of Strategy Set and Experiment

The final purpose of strategy evaluation is to judge the quality of strategy. Good strategies could achieve good results. By the directed graph it is possible

Table 2. States of Nodes in RedI Strategy Subset TA

No.	State	No.	State	No.	State
01	begin	06	31334243	11	42435152
02	31344243	07	32334243	12	42525253
03	41435253	08	23313243	13	goal
04	42435253	09	33414253		
05	32435253	10	41435152		

Table 3. States of Nodes in RedII Strategy Subset TA

No.	State	No.	State	No.	State
01	begin	06	23313243	11	33414253
02	31344243	07	41435253	12	41435152
03	32334243	08	42435253	13	42435152
04	22324243	09	42525253	14	goal
05	31334243	10	32435253		

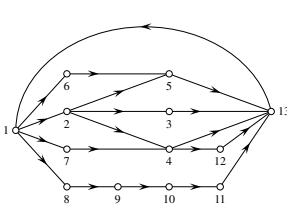


Fig. 5. Digraph of RedI Subset TA

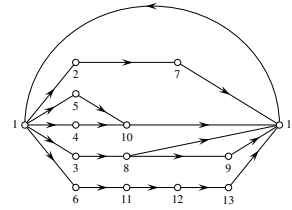


Fig. 6. Digraph of RedII Subset TA

to compute the probability of goal. Therefore, the probability of goal becomes a benchmark of strategy set evaluation. Better strategy set would achieve higher probability of goal score.

There is a strategy set of Red team called RedI. In order to compute the goal probability, we abstract the nodes in RedI subset TA, as shown in Table 2.

Fig. 5 displays the complete digraph of RedI strategy subset TA, where node 1 is the node of game beginning, and node 13 is goal score. The node 13 contains the states of ‘52536263’, ‘42536263’, ‘52525362’, and ‘43536263’. According to Fig. 5, we have the following equations.

$$P(1) = P(1|13)P(13) \tag{9}$$

$$P(2) = P(2|1)P(1) \tag{10}$$

$$P(3) = P(3|2)P(2) \tag{11}$$

$$P(4) = P(4|2)P(2) + P(4|7)P(7) \tag{12}$$

$$P(5) = P(5|2)P(2) + P(5|6)P(6) \tag{13}$$

$$P(6) = P(6|1)P(1) \tag{14}$$

$$P(7) = P(7|1)P(1) \tag{15}$$

$$P(8) = P(8|1)P(1) \tag{16}$$

$$P(9) = P(9|8)P(8) \tag{17}$$

$$P(10) = P(10|9)P(9) \tag{18}$$

$$P(11) = P(11|10)P(10) \tag{19}$$

$$P(12) = P(12|4)P(4) \tag{20}$$

$$P(13) = P(13|5)P(5) + P(13|3)P(3) + P(13|4)P(4) + P(13|12)P(12) + P(13|11)P(11) \tag{21}$$

Then the matrix A is

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \frac{1}{2} & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}, \tag{22}$$

where we select rules stochastically at the node 1, node 2 and node 4. The eigenvector with eigenvalue 1 for matrix A is

$$x = [0.6092 \quad 0.1523 \quad 0.0508 \quad 0.2031 \quad 0.2031 \quad 0.1523 \quad 0.1523 \quad 0.1523 \quad 0.1523 \quad 0.1523 \quad 0.1015 \quad 0.6092 \quad]^T. \tag{23}$$

Therefore, the probability of goal score in RedI strategy subset TA is

$$P(13) = 0.6092. \tag{24}$$

Now we would like to improve the strategy set of Red team. After modifying some rules, we get a new strategy set RedII. The abstracted nodes in RedII subset TA is shown in Table 3, then we can draw the complete directed graph of RedII strategy subset TA (see Fig. 6).

Fig. 6 displays the complete digraph of RedII strategy subset TA, where node 1 is the node of game beginning, and node 14 is goal score. The node 14 contains the states of ‘52536263’, ‘42536263’, ‘33536263’, ‘52525362’, and ‘43536263’. Similarly, we have

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \text{c} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \text{c} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \text{c} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \text{c} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \text{c} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \text{c} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \frac{1}{2} & 1 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}, \tag{25}$$

where we select rules stochastically at the node 1 and node 8. The eigenvector with eigenvalue 1 for matrix A is

$$x = [0.6238 \quad 0.1248 \quad 0.1248 \quad 0.1248 \quad 0.1248 \quad 0.1248 \quad 0.1248 \quad 0.1248 \quad 0.1248 \quad 0.0624 \quad 0.2495 \quad 0.1248 \quad 0.1248 \quad 0.1248 \quad 0.6238]^T. \tag{26}$$

Therefore, the probability of goal score in RedII strategy subset TA is

$$P(14) = 0.6238. \tag{27}$$

Apparently, $P(14)$ of RedII is greater than $P(13)$ of RedI, that means RedII is better than RedI.

The idea of strategy set evaluation can be testified by simulated experiment. Robot soccer simulator is a good test engine for strategy set. In the simulator, two teams will share the same tactics, which would eliminate the impact of simulator, and distinctly display differences between the two strategy sets. Therefore, in order to get rid of the influence of tactics, we carry out all experiments on the platform of simulator.

Now we have another team called Blue team. Due to space limitation, we do not list the strategy set of Blue team. Table 4 displays ten games' results of RedI vs Blue, where RedI team got three wins, three defeats and four ties. Table 5 lists ten games' results of RedII vs Blue, where RedII team got four wins, two defeats and four ties. Obviously, the strategy set of RedII is better than RedI, which is coincident with the results of Eq. (24) and Eq. (27). Therefore, the presented method of strategy set evaluation is validated. In addition, the value of goal score probability in strategy set Blue is same to RedI, i.e. 0.6092, which

Table 4. RedI vs Blue

Team	Score										
<i>RedI</i>	1	0	1	2	0	0	1	2	3	3	1
<i>Blue</i>	1	0	1	0	1	0	0	0	0	4	3

Table 5. RedII vs Blue

<i>Team</i>	<i>Score</i>									
<i>RedII</i>	1	0	1	2	1	1	1	0	2	0
<i>Blue</i>	0	1	0	3	1	1	0	0	1	0

explains the reason why RedI got three wins, three defeats and four ties. This validates our method from the other side.

5 Conclusion and Future Work

Strategy evaluation in robot soccer is a very important issue in the field of multi-robot coordinated control system. In this work, we present a graph theory-based method to evaluate the strategy set in robot soccer. Firstly, we present the description of strategy, make clear the morphology of strategy set, depict the four types of strategy subset. Secondly, we present the model of strategy set evaluation, in which the strategy set could be evaluated by probability of goal score, because better strategy set would achieve higher probability of goal score. Thirdly, we illustrate an example to compute the goal score probability of RedI and RedII respectively. The value of RedII is greater than that of RedI, it means strategy set RedII is better than RedI. Finally, in order to testify our method, twenty simulation games are played where ten games for RedI vs Blue and another ten games for RedII vs Blue, so that we can observe the performance difference between RedI and RedII. In fact, the results of simulation games are coincident with our calculation, which validates our method.

As future work, we are interested in carrying more experiments with bigger strategy sets. In our method, the probability of goal score is the benchmark of strategy set evaluation, which is the corresponding component of coefficient matrix eigenvector. Apparently, the eigenvector is related to the topology of directed graph. Therefore in future we would like to seek an optimized directed graph structure of strategy set.

References

- [1] Bowling, M., Browning, B., Chang, A., Veloso, M.: Plays as team plans for coordination and adaptation. In: Polani, D., Browning, B., Bonarini, A., Yoshida, K. (eds.) RoboCup 2003. LNCS (LNAI), vol. 3020, pp. 686–693. Springer, Heidelberg (2004)
- [2] Bowling, M., Browning, B., Veloso, M.: Plays as effective multiagent plans enabling opponent-adaptive play selection. In: International Conference on Automated Planning and Scheduling, ICAPS 2004, Whistler, Canada, vol. 1, pp. 376–383. AAAI Press (June 2004)
- [3] Browning, B., Bruce, J., Bowling, M., Veloso, M.: Stp: skills, tactics, and plays for multi-robot control in adversarial environments. *Journal of Systems and Control Engineering* 219(1), 33–52 (2005)

- [4] Chapman, S., Derse, E., Hansen, J.: Soccer Coaching Manual. LA84 Foundation, Los Angeles (2007)
- [5] Dylla, F., Ferrein, A., Lakemeyer, G.: Acting and deliberating using golog in robotic soccer – a hybrid architecture. In: International Cognitive Robotics Workshop, CogRob 2002, Edmonton, Canada. AAAI Press (July 2002)
- [6] Dylla, F., Ferrein, A., Lakemeyer, G., Murray, J., Obst, O., Röfer, T., Schiffer, S., Stolzenburg, F., Visser, U., Wagner, T.: Approaching a formal soccer theory from behaviour specifications in robotic soccer, ch. 6. Computers in Sport, pp. 161–185. WIT Press, Southampton (2008)
- [7] Dylla, F., Ferrein, A., Lakemeyer, G., Murray, J., Obst, O., Röfer, T., Stolzenburg, F., Visser, U., Wagner, T.: Towards a league-independent qualitative soccer theory for robocup. In: Nardi, D., Riedmiller, M., Sammut, C., Santos-Victor, J. (eds.) RoboCup 2004. LNCS (LNAI), vol. 3276, pp. 611–618. Springer, Heidelberg (2005)
- [8] Floyd, M.W., Esfandiari, B., Lam, K.: A case-based reasoning approach to imitating robocup players. In: International Florida Artificial Intelligence Research Society Conference, FLAIRS 2008, Florida, USA, pp. 251–256. AAAI Press (May 2008)
- [9] Huang, Z., Yang, Y., Chen, X.: An approach to plan recognition and retrieval for multi-agent systems. In: First RoboCup Australian Open Workshop on Adaptability in Multi-Agent Systems. Citeseer (2003)
- [10] Jolly, K.G., Ravindran, K.P., Vijayakumar, R., Sreerama Kumar, R.: Intelligent decision making in multi-agent robot soccer system through compounded artificial neural networks. In: Robotics and Autonomous Systems, vol. 55(7), pp. 589–596 (2007)
- [11] Kleiner, A., Dietl, M., Nebel, B.: Towards a life-long learning soccer agent. In: Kaminka, G.A., Lima, P.U., Rojas, R. (eds.) RoboCup 2002. LNCS (LNAI), vol. 2752, pp. 126–134. Springer, Heidelberg (2003)
- [12] Lam, K., Esfandiari, B., Tudino, D.: A scene-based imitation framework for robocup clients. In: AAAI Workshop on Modeling Other Agents from Observations (2006)
- [13] Lattner, A.D., Miene, A., Visser, U., Herzog, O.: Sequential pattern mining for situation and behavior prediction in simulated robotic soccer. In: Bredenfeld, A., Jacoff, A., Noda, I., Takahashi, Y. (eds.) RoboCup 2005. LNCS (LNAI), vol. 4020, pp. 118–129. Springer, Heidelberg (2006)
- [14] Lee, J., Ji, D., Lee, W., Kang, G., Joo, M.: A tactics for robot soccer with fuzzy logic mediator. In: Hao, Y., Liu, J., Wang, Y.-P., Cheung, Y.-m., Yin, H., Jiao, L., Ma, J., Jiao, Y.-C. (eds.) CIS 2005. LNCS (LNAI), vol. 3801, pp. 127–132. Springer, Heidelberg (2005)
- [15] Lucchesi, M.: Coaching the 3-4-1-2 and 4-2-3-1. Reedswain (August 2002)
- [16] Miene, A., Visser, U., Herzog, O.: Recognition and prediction of motion situations based on a qualitative motion description. In: Polani, D., Browning, B., Bonarini, A., Yoshida, K. (eds.) RoboCup 2003. LNCS (LNAI), vol. 3020, pp. 77–88. Springer, Heidelberg (2004)
- [17] Nakashima, T., Takatani, M., Namikawa, N., Ishibuchi, H., Nii, M.: Robust evaluation of robocup soccer strategies by using match history. In: IEEE Congress on Evolutionary Computation, CEC 2006, Vancouver, BC, Canada, pp. 1195–1201. IEEE (July 2006)
- [18] Nakashima, T., Takatani, M., Udo, M., Ishibuchi, H.: An evolutionary approach for strategy learning in robocup soccer. In: IEEE International Conference on Systems, Man and Cybernetics, SMC 2004, Hague, Netherlands, vol. 2, pp. 2023–2028. IEEE (October 2004)

- [19] Nakashima, T., Takatani, M., Udo, M., Ishibuchi, H., Nii, M.: Performance evaluation of an evolutionary method for robocup soccer strategies. In: Bredenfeld, A., Jacoff, A., Noda, I., Takahashi, Y. (eds.) RoboCup 2005. LNCS (LNAI), vol. 4020, pp. 616–623. Springer, Heidelberg (2006)
- [20] Park, J.H., Stonier, D., Kim, J.H., Ahn, B.H., Jeon, M.G.: Recombinant rule selection in evolutionary algorithm for fuzzy path planner of robot soccer. In: Freksa, C., Kohlhase, M., Schill, K. (eds.) KI 2006. LNCS (LNAI), vol. 4314, pp. 317–330. Springer, Heidelberg (2007)
- [21] Riedmiller, M., Merke, A., Meier, D., Hoffmann, A., Sinner, A., Thate, O., Ehrmann, R.: Karlsruhe brainstormers - a reinforcement learning approach to robotic soccer. In: Stone, P., Balch, T., Kraetzschmar, G.K. (eds.) RoboCup 2000. LNCS (LNAI), vol. 2019, pp. 367–372. Springer, Heidelberg (2001)
- [22] Ros, R., Arcos, J.L.: Acquiring a robust case base for the robot soccer domain. In: International Joint Conference on Artificial Intelligence, IJCAI 2007, Hyderabad, India, pp. 1029–1034. AAAI Press (January 2007)
- [23] Ros, R., Arcos, J.L., Lopez de Mantaras, R., Veloso, M.: A case-based approach for coordinated action selection in robot soccer. *Artificial Intelligence* 173(9-10), 1014–1039 (2009)
- [24] Ros, R., Veloso, M.: Executing multi-robot cases through a single coordinator. In: International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2007, Honolulu, Hawaii, pp. 1272–1274. ACM (May 2007)
- [25] Wu, C.J., Lee, T.L.: A fuzzy mechanism for action selection of soccer robots. *Journal of Intelligent and Robotic Systems* 39(1), 57–70 (2004)

Spatial and Frequency Domain–Based Feature Fusion Method for Texture Retrieval

Rurui Zhou

School of Computer Science, Yangtze University
Zhrr@yangtzeu.edu.cn

Abstract. This work presents a novel feature fusion method for texture retrieval. Considering the advantages of both the spatial and frequency domain, we first carry on the experiments in spatial domain and frequency domain respectively. On one hand, sober and histogram feature are used to calculate the similarity. On the other hand, Fourier is applied to obtain the frequency feature. Then a feature fusion scheme is used to join the two features came from spatial and frequency domain. Experimental results on MIT texture database show that the proposed method is effective.

Keywords: texture retrieval, spatial domain, frequency domain, fusion.

1 Introduction

Texture has played an important role in computer vision. It is almost everywhere around our daily life. This vital property provides us abundant information for feature extraction. The texture consists of varieties of types. For example, a smooth texture appears in the furniture demonstrating its good looking for appreciation, the hardness texture in the wildness depicting the nature of the life. In addition, some of the texture have fix pattern and some have dynamic pattern such as the wave and the swaying trees. However, due to the strong intrinsic of the texture between local area and the difference between different textures, we benefit a lot from these advantages. People also make good use of the texture information to recognize all of the things around us.

The statistic-based method has been developed for many years. Many researches validate that this type of method is one of the effective one, among which the grey level co-occurrence matrix (GLCM) is good texture analysis tool that has been a mainstream method. Haralick et al. [1] first proposed the GLCM method when he was doing the research on land use. The GLCM method describes the appearance probability that belongs to a couple of grey level. Although GLCM method owns good discrimination ability, it still suffers from high computation complexity, especially for the texture classification on pixel level. The researchers have tried many methods to improve the original version of GLCM. Soh and Tsatsoulis [2] reduced the computation burden of GLCM through different scale and direction. Ulaby et al. [3] discovered that four features that in terms of contrast, inverse gap, correlation and energy is irrelevant, which is convenient for computation and gives high classification accuracy. Baraldi [4] had made a detail research on the six high

texture features and recognized that the contrast and the entropy are the two most important features. B. Hua [5] had carried on the optimization of computation of GLCM and had obtained three irrelevant features with best discrimination ability, namely the contrast, entropy and correlation. Although the above mentioned methods can reduce the computation burden but it cannot solve the existing problem of GLCM. Aiming at tackling with the above problems, Clausi et al. [6] made a deep inside of GLCM and finally obtained a good improvement of the original version. In addition, Walker et al. [7] proposed an adaptive multiple scale GLCM method and an improved version of GLCM based on genetic algorithms. Experimental results show that the classification error rates of both proposed method are obviously lower than traditional GLCM method and reduce the computation burden of feature selection. Kandaswamy [8] analyzed the computation complexity of GLCM. Enlightened by the statistic model, they proposed an efficient texture analysis method. Their conclusion shows that the analogy texture feature is able to improve the efficiency of texture analysis, but they do not arouse the decreasing of the classification error rate.

The parameter-based method first built a model of the image texture, and then regarded the texture feature extraction as the process of parameter evaluation. It is an essential work to evaluate the optimal parameters. The random field model tries to use the probability model to depict the random process. The essence of the random field model is to depict the statistical dependence relation, among which Markov random field (MRF) model is the most popular one. The basic idea of MRF texture modeling is to describe the statistical feature of the texture according to the conditional probability distribution of a pixel in terms of its neighbor pixel. According to the Hammersley-Clifford theory [9], there is an equivalent property between the random field of MRF and Gibbs. Hence, it is convenient to depict the space constrains of the images through conditional probability function or the union distribution. Besides the MRF model, McCormick et al. [10] proposed to use the autoregressive model for texture analysis. This method evaluates the pixel level from the corresponding pixel level of its neighborhood, where the evaluation of the parameters are carried on by using the criterion of minimum mean square error or maximum likelihood parameter estimation. There is a notable change to the parameters of the model for the smooth texture, but no changes for the rough texture.

In this paper, we present a novel feature fusion method for texture retrieval. Considering the advantages of both the spatial and frequency domain, we first carry on the experiments in spatial domain and frequency domain respectively. On one hand, sobel and histogram feature are used to calculate the similarity. On the other hand, Fourier is applied to obtain the frequency feature. Then a feature fusion scheme is used to join the two features came from spatial and frequency domain. Experimental results on MIT texture database show that the proposed method is effective.

The paper is organized as follows. In the next section we introduce related work. Section 3 describes our method. In section 4, we give a number of experiment results. Section 5 offers our conclusions.

2 Related Work

The definition of texture can be stated as follow: The pixel arrangement of the surface of an image, which describes the main property of the image and enables us to tell

apart, two different types of things. With the good merits of the texture, there has been a wide scope of applications in a variety of fields. Texture analysis has made great contribution in remote image, X-ray photo, cell image processing. It also can be recognized as a tool that can be used for depicting any arrangement of the substance such as the lung and vascular texture in medical X-ray photo, the texture of the rocks in aerospace terrain. In addition, texture analysis has also been concerned in various biometrics recognition systems, such as the iris recognition system, vein recognition system, palm print recognition system, face recognition and so on. Even though such tool has been used for many fields for a long time, it still suffers from some difficulties which can be stated as follows: (1) It is generally very hard to find a proper mathematical model to describe every detail of the texture; (2) There is still some randomness, namely irregular patterns, in the texture which is hard to be well copied with. (3) It is normally not a realistic work to exploit only the local or the global feature for texture classification. That is to say, a single texture detection method cannot obtain a much better result. (4) It is a challenge work to look for the robust and efficient feature for texture analysis. This hardness on one hand restricts the effectiveness of the current applications in terms of texture analyze, on the other hand facilitate the researchers to make an improvement of the related work.

The frequency-based method assumes that the distribution of the energy in frequency domain can be used to identify the texture. It transforms the texture from one space to another space through a certain linear transform, filters or filter banks, and then applied a certain energy-based criterion to extract the texture feature. The discrete cosine transform [11] and the fourier transform [12] are commonly used methods for feature extraction in frequency domain. Gabor filter is also a good tool for feature extraction. Its main idea is that: Different texture usually owns different central frequency and bandwidth. Hence, the Gabor filter banks can be designed for feature extraction according to such frequency and bandwidth. For each Gabor filter, only the texture that has the same frequency as the Gabor filter can go through the filter, while the texture with other frequency cannot pass. Therefore, the prior task for Gabor filter design is to consider the design of the parameters of each filter and the layout of the Gabor filter banks. Dunn and Higgins [13] had made a fundamental work to demonstrate the detail design criterion of single Gabor filter. The Gabor filter banks should cover the whole space. With this reason, the parameters of frequency, scale and position should be properly set so as to prevent the overlap of frequency banks on the same radius. The frequency bank in radial direction with different radius should not be overlapped too.

The representation-based method holds that the texture consists of basis patches or lied in a lower subspace. Hence, the basis patches can be used for constructing the texture. The syntactic texture description method [14] assumes that the description of the texture of a class can form a language that is expressed by the corresponding grammar. The mathematical morphology method [15] used structural element to look for the repetition property in the space. When the binary texture image is eroded by the structural element, the texture property will appear on the appearance of the eroded image. In addition, another group of representation methods represent the texture in a lower subspace. These methods include PCA, SRC and so on.

3 The Proposed Method

In this section, we propose a fusion method to achieve better texture retrieval result. First, we carry on the experiments in both spatial and frequency domain. Then, the proposed fusion method is used to retrieve the images.

A. The spatial filter

- Sobel operator

For the function $f(x, y)$ of an image, the corresponding gradient is defined by using the expression (1).

$$\nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \tag{1}$$

The module of ∇f is given by Equation (2)

$$\nabla f = mag(\nabla f) = [G_x^2 + G_y^2]^{\frac{1}{2}} = \left[\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2 \right]^{\frac{1}{2}} \tag{2}$$

However, it costs lots of computation as soon as Equation (2) is used for calculation, since it contains the square operations and a root operation. In real application, we use the absolute value operation to replace the square and root operations. Therefore, Equation (3) is used to calculate the gradient of function $f(x, y)$. In addition, we use the difference operation over the image to approximate the calculation of Equation (3). With this reason, two masks along X direction and Y direction are used for the gradient calculation. We regard the two masks as sobel operators, which is shown in Figure 1.

$$\nabla f \approx |G_x| + |G_y| \tag{3}$$



Fig. 1. Sobel operator. (a) is the operator along X direction while (b) is the operator along Y direction

¹ Identify applicable sponsor/s here. (sponsors).

- **Histogram feature**

The object can be described by histogram feature $P = (p_1, p_2, \dots, p_m)$, among which

$$p_u = C \sum_{i=1}^n \delta[c(x_i) - u], (u = 1, 2, \dots, m) \quad (4)$$

$$C = \frac{1}{\sum_{u=1}^m \sum_{i=1}^n \delta[c(x_i) - u]} \quad (5)$$

$$\delta(x) = \begin{cases} 1, & \text{if } x = 0 \\ 0, & \text{if } x \neq 0 \end{cases} \quad (6)$$

$$c(x_i) = \left[\frac{H(x_i)}{m} \right] + 1 \quad (7)$$

Where x_i is the pixel value, $H(x_i)$ is the i th pixel value in gray channel, n is the pixel number and m is the bin number.

B. The frequency filter

- **Two-dimensional discrete Fourier transform**

For an image with dimension $M \times N$, the discrete Fourier transform of its function $f(x, y)$ can be expressed as Equation (8)

$$F(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-j2\pi(ux/M + vy/N)} \quad (8)$$

Where $u = 0, 1, \dots, M-1$, $v = 0, 1, \dots, N-1$, $x = 0, 1, \dots, M-1$, $y = 0, 1, \dots, N-1$. The spectrum and the phase angle are shown in Equation (9) and (10), respectively.

$$|F(u, v)| = \left[R^2(x, y) + I^2(x, y) \right]^{\frac{1}{2}} \quad (9)$$

$$\phi(u, v) = \arctan \left[\frac{I(u, v)}{R(u, v)} \right] \quad (10)$$

Usually, we conduct the discrete Fourier transform on function $f(x, y)(-1)^{x+y}$. According to the property of exponentiation, we have

$$\zeta[f(x, u)(-1)^{x+y}] = F(u - M/2, v - N/2) \tag{11}$$

Where $\zeta[\bullet]$ stands for the Fourier transform. Equation (11) transfer the center of $F(u, v)$ to $(M/2, N/2)$ under frequency coordinate.

C. The fusion scheme

Since the spatial and frequency domain have its own advantages in texture retrieval. In this work, we fuse the results of the methods came from spatial and frequency domain respectively. See in Fig. 1, as soon as we obtain the histogram feature and the frequency feature, we use Equation (12) to obtain the final fused feature.

$$S = \alpha A + (1 - \alpha)B \tag{12}$$

Where A is the histogram feature, B is the frequency feature, S is the final fused feature and α is an adjusted factor ranged from 0 to 1.

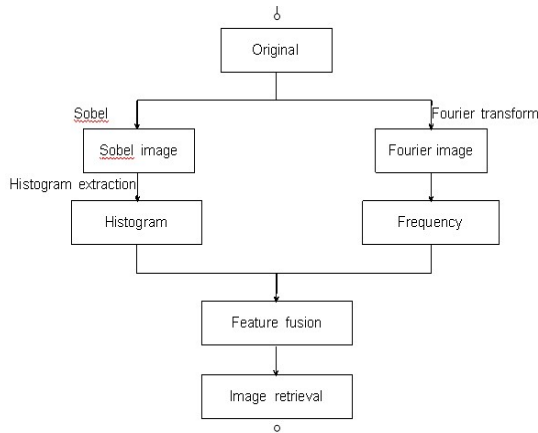


Fig. 2. Fusion scheme of texture retrieval

4 Experimental Results

To validate the effectiveness of the proposed fusion method, we carry on the experiments on MIT texture database. The database has 40 categories. For each category, it contains only one image. We divide the image into four parts. Hence, we obtained 160 images, with 40 categories. For each category, there are only four samples. Figure 3 shows some sample images that we use in the experiments. To test

the precision of the proposed method, for a testing image, the retrieval system return four images which are probably the same class as the testing image. The number of the returned images which are the same class as the testing image will be accumulated and finally divide the total number of the returned images so as to obtain the precision of the retrieval system.

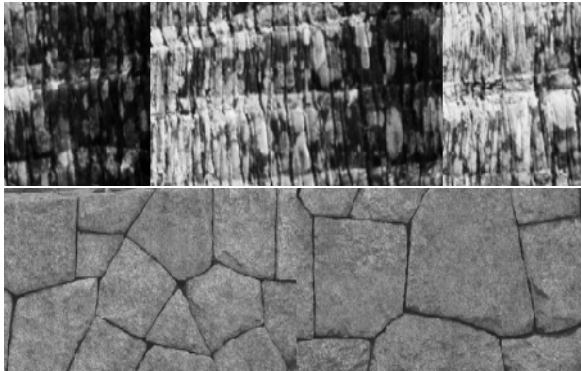


Fig. 3. Sample images from MIT texture database

We run the sobel-histogram method in spatial domain, frequency-based method and the fusion method on the obtained database. Figure 4 shows the experimental results. With the increasing of the alpha factor, the result of the fusion method is becoming much better than its original version.

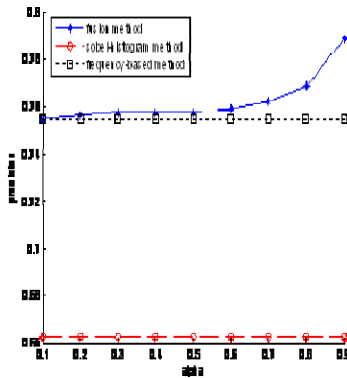


Fig. 4. Precision of the proposed method

We also run our experiments under different scheme whose experimental results are shown in Table 1. The experimental result shows that the proposed fusion scheme shows a competitive result than other methods.

Table 1. Precision under different methods

Method	precision
NN [16]	33.59%
Sobel+NN	28.91%
Sobel+Partitions+Histogram+NN	65.78%
partitions+Fourier+NN	65.78%
Fourier+Histogram+NN	36.72%
Sobel+Fourier+NN	5.15%
Sobel+Histogram+NN	66.25%
Fourier+NN	75.47%
fusion method	78.91%

5 Conclusions

In this paper, we have used a fusion method which takes advantage of spatial and frequency domain. In spatial domain, we first extract the sobel image of the testing image, then the histogram feature is obtained. In frequency domain, the Fourier transform is performed to extract the frequency feature. Finally, we combine the obtained two features for texture retrieval. Experimental results on MIT texture database show that, the proposed method is competitive and effective.

References

1. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. *IEEE Trans. on Systems, Man, and Cybernetics* 3(6), 610–621 (1973)
2. Soh, K.S., Tsatsoulis, C.: Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *IEEE Trans. on Geoscience and Remote Sensing* 37(2), 780–795 (1999); Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. *ACM Computing Survey* 34(4), 399–485 (2003)
3. Ulaby, F.T., Kouyate, F., Brisco, B., et al.: Textural information in SAR images. *IEEE Transactions on Geoscience and Remote Sensing* 24(2), 235–245 (1986); Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*, 2nd edn. Prentice-Hall, Upper Saddle River (2002)
4. Baraldi, A., Parmiggiani, F.: An investigation of the textual characteristics associated with gray level co-occurrence matrix statistical parameters. *IEEE Trans. on Geoscience and Remote Sensing* 33(2), 293–304 (1995)
5. Hua, B., Ma, F.-L., Jiao, L.-C.: Research on computation of GLCM of image texture. *Acta Electronica Sinica* 34(1), 155–158 (2006)
6. Clausi, D.A., Jernigan, M.E.: A fast method to determine co-occurrence texture features. *IEEE Trans. on Geoscience and Remote Sensing* 36(1), 298–300 (1998)

7. Walker, R.F., Jackway, P.T., Longstaff, I.D.: Recent developments in the use of co-occurrence matrix for texture recognition. In: Proceedings of IEEE Conference on Digital Signal Processing, Santorini, Greece, vol. 1, pp. 63–65 (1997)
8. Kandaswamy, U., Adjeroh, D.A., Lee, M.C.: Efficient texture analysis of SAR imagery. *IEEE Trans. on Geoscience and Remote Sensing* 43(9), 2075–2083 (2005)
9. German, S., German, D.: Stochastic relaxation, gibbs distribution and Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 6(6), 721–741 (1984)
10. McCormick, B.H., Jayaramamurthy, S.N.: Time series model for texture synthesis. *International Journal of Computer Information Science* 3(4), 329–343 (1974)
11. Ng, I., Tan, T., Kittler, J.: On local linear transform and Gabor filter representation of texture. In: Proceeding of the 11th IAPR Conference on Image, Speech and Signal Analysis, Hague, Netherlands, pp. 627–631 (1992)
12. Zhou, F., Feng, J., Shi, Q.: Image segmentation based on local fourier transform. In: Proceedings of International Conference on image Processing, Wuhan, China, pp. 610–613 (2001)
13. Dunn, D., Higgins, W.E., Wakeley, J.: Texture segmentation using 2-D Gabor elementary functions. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 16(2), 130–149 (1994)
14. Pavlidis, T.: *Structural Descriptions and Graph Grammars*, pp. 86–103. Springer, Berlin (1980)
15. Soille, P.: *Morphological Image Analysis: Principles and Applications*, pp. 289–317. Springer Press, Berlin (2003)
16. Cover, T., Hart, P.: Nearest neighbor pattern classification. *IEEE Transactions in Information Theory* IT-13, 21–27 (1967)

Comparisons of Typical Discrete Logistic Map and Henon Map

Bingbing Song and Qun Ding*

Electronic Engineering College, Heilongjiang University,
Harbin, China
songbingbing1988@163.com, qunding@aliyun.com

Abstract. Applying chaos theory to the encryption scheme has become a hot spot. Although lots of chaotic maps have been proposed, they don't have advantages in all respects. In this paper, the typical one-dimensional Logistic map and two-dimensional Henon map are studied. The digital output sequences of Logistic map and Henon map can be generated by building their models on DSP Builder platform. And the sequences are tested and compared according to the statistical properties, including balance test, run test and autocorrelation test. Meanwhile, run the transformed VHDL projects in Quartus II environment to compare these two kinds of resource utilizations. The results show that digital Henon sequences have better pseudo-randomness, but Logistic project uses less hardware resources.

Keywords: chaotic maps, DSP Builder, statistical properties, hardware resources.

1 Introduction

Since the famous American meteorologist Lorenz discovered chaos in 1963, more and more scholars have devoted themselves to studying chaos theory. Chaos is a kind of complex movement [1]. Because of basic features of chaotic sequences, such as good pseudo-randomness, complexity, initial value sensitivity, etc., thus, they have good prospects in the field of cryptography. In recent years, using chaotic systems in cryptography has become a trend, especially in sequential cipher [2]. Nowadays, researchers often focus on improving chaotic sequences or putting forward new encryption methods. Yunpeng Zhang raised a new idea that the effect of short period of Logistic could be overcome by using the improved Chebyshev system to disturb the chaotic parameter μ of Logistic [3]. Qun Ding achieved the hardware circuits of Logistic chaotic system and Lorenz chaotic system, and they were applied to network data encryption and serial data encryption devices [4]. Chenhui Jin proposed to attack chaotic stream cipher by

* This paper is supported by Innovated Team Project of 'Modern Sensing Technology' in colleges and universities of Heilongjiang Province (No. 2012TD007) and Institutions of Higher Learning by the Specialized Research Fund for the Doctoral Degree (No.20132301110004).

combining the ideas of linear cryptanalysis with the ways of attacking chaotic ciphers [5,6]. However, few people make a comparative study of the performances of different chaotic systems. So the analyses are discussed in this paper.

States of chaotic systems usually can be described by non-linear differential equations or difference equations, and they are commonly referred to as continuous chaos and discrete chaos. Typical continuous chaotic systems include Lorenz system, Rossler system and Duffing system [7,8,9], and typical discrete chaotic systems are Logistic system, Tent system and Henon system [10,11,12]. In this paper, the Logistic system and Henon system are studied, and their digital output sequences are also compared. The detail of the paper is organized as follow: firstly, the two models of Logistic map and Henon map have been built by the DSP Builder modules in Simulink library based on their equations; secondly, the digital output sequences of these two models are compared by balance test, run test and autocorrelation test. In addition, the hardware resources these two models use are analyzed at the same time; finally, the results show that digital Henon sequences have better pseudo-randomness, but Logistic project uses less hardware resources.

2 Digital Sequences Generation of Chaotic Map

2.1 Chaotic Mathematical Equation

Logistic Map. Logistic map is a simple and widely studied dynamic system. It is defined as follows [13]:

$$x_{n+1} = f(x_n) = \mu x_n(1 - x_n) \quad (1)$$

where μ is the bifurcation parameter, n is the n th iteration, and x_n is the n th state. The parameter μ is very important. It determines whether the system is chaotic or not.

Currently, chaos theorists have proposed some tools, such as Lyapunov exponent, bifurcation diagram, and phase diagram [14,15]. Fig. 1 and Fig. 2 present the bifurcation diagram and the Lyapunov exponent of Logistic map respectively.

From Fig. 1 and Fig. 2 we can find that as the parameter μ increases, the system gradually becomes chaotic [16]. And when $\mu \in (3.5, 4]$, the system is chaotic.

Henon Map. Inspired by the studies of globular clusters and Lorenz system, the French astronomer Henon put forward Henon map [17]. It is defined as follows [18]:

$$\begin{cases} x_{n+1} = 1 - ax_n^2 + y_n \\ y_{n+1} = bx_n \end{cases} \quad (2)$$

where a and b are parameters, n is the n th iteration, x_n and y_n are the n th states. Henon map is the simplest two-dimensional map. Fig. 3 and Fig. 4 respectively present the bifurcation diagram and the largest Lyapunov exponent of Henon map for $b = 0.3$ and $a \in [0, 1.4]$.

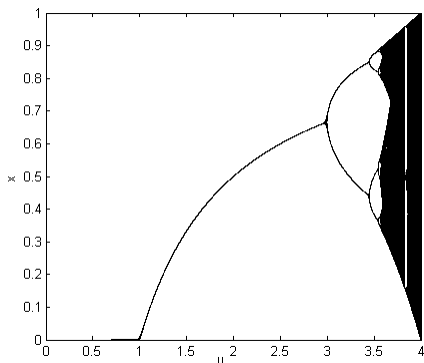


Fig. 1. Bifurcation diagram of Logistic map

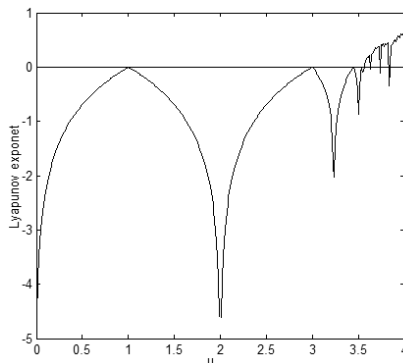


Fig. 2. Lyapunov exponent of Logistic map

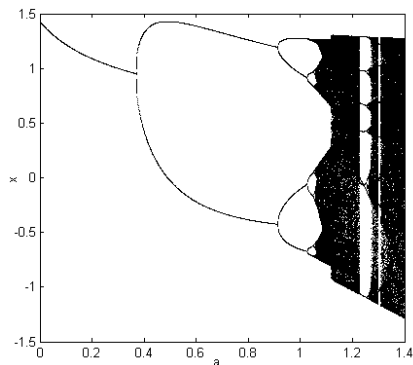


Fig. 3. Bifurcation diagram of Henon map

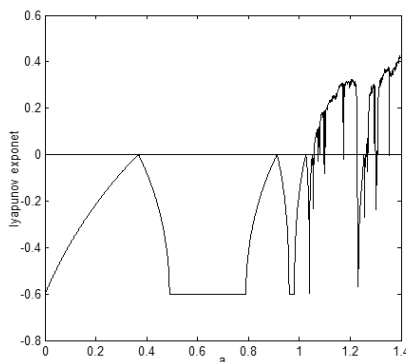


Fig. 4. Largest Lyapunov exponent of Henon map

Fig. 3 and Fig. 4 show that as the parameter a increases, the system gradually becomes chaotic. When the largest exponent is larger than zero, the system is chaotic.

2.2 DSP Builder Model

According to chaotic mathematical equations, the two models of Logistic system and Henon system are built by using DSP Builder modules in Simulink library. And then the digital output sequences are obtained.

Logistic Model. According to Eq. 1, the Logistic model consists of initial value module, data selector module, delay unit module, multiplication module, addition operation module, fixed-point to floating-point conversion module and quantization module. All the structures diagram of Logistic map is shown in Fig. 5.

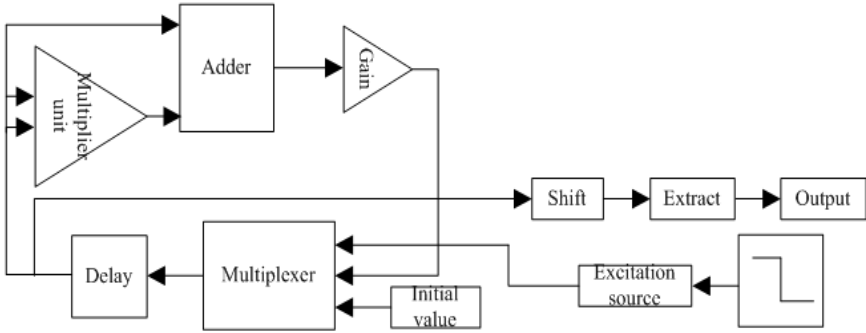


Fig. 5. Structures diagram of Logistic map

In Fig. 5, the implementation of the whole process consists of some steps: firstly, when the excitation source is high, the initial value will enter the multiplexer module, otherwise, the whole iteration starts, including delay, multiplier, gain, and adder; eventually, after shift and extraction, the binary output sequences are obtained from the output module.

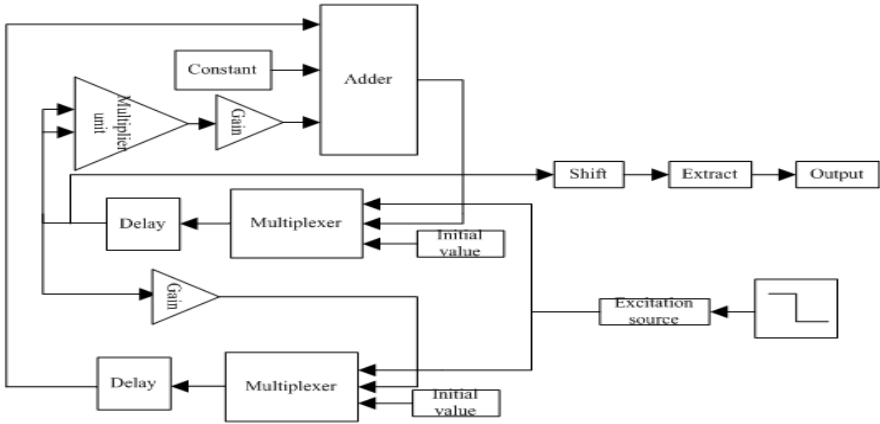


Fig. 6. Structures diagram of Henon map

Henon Model. According to Eq. 2, the Henon model is built. Fig. 6 shows the structures diagram of Henon map.

In Fig. 6, initial value 1 module is the initial value of x , and initial value 2 module is the initial value of y . The implementation of the whole process is similar to that of Logistic.

2.3 Simulink Simulation

After constructing the chaotic models, the next step is to verify. Simulink can realize the simulation function. Before starting the simulation, it is necessary to set the simulation time, the step interval, etc., and oscilloscope modules are also needed to observe the outputs. The simulation results of Logistic model and Henon model are shown in Fig.7 and Fig.8 respectively.

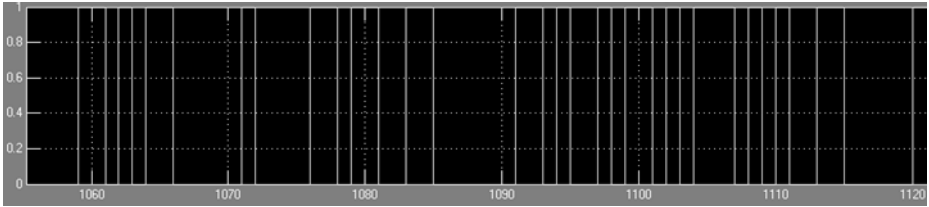


Fig. 7. Simulation result of Logistic model

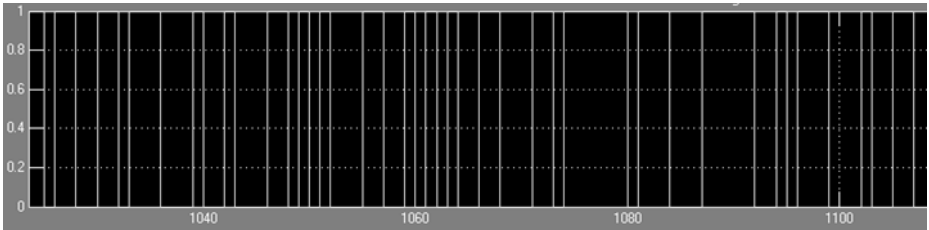


Fig. 8. Simulation result of Henon model

From Fig.7 and Fig.8, we can draw a conclusion that both of the two chaotic output sequences are 0-1 binary sequences and seem to be non-periodic.

3 Comparisons of Digital Output Sequences

After obtaining the digital output sequences, the comparisons are made according to the statistical properties, including balance test, run test and autocorrelation test. Thus, we can determine which chaotic map has better pseudo-randomness. The experimental tests are shown below in details.

3.1 Balance Test

Balance test is used to test whether the number of 0 and 1 in the binary sequence is approximately equal or not.

Unbalance formula is defined as follows [19]:

$$E(N) = \frac{|Q_1 - Q_0|}{N} \% \tag{3}$$

Table 1. Balance test of of digital sequences of Logistic map and Henon map

Length	Chaotic map	Number of 0	Number of 1	Degree of unbalance
1000000	Logistic	497187	502813	0.0056
	Henon	501908	498092	0.0038
100000	Logistic	50116	49884	0.0023
	Henon	50444	49556	0.0089
10000	Logistic	4953	5047	0.0094
	Henon	5029	4971	0.0058
1000	Logistic	495	505	0.0100
	Henon	516	484	0.0320

where Q_1 and Q_0 represent the number of 1 and the number of 0 respectively, and N is the length of the sequence. According to Eq. 3, balance test results of digital sequences of Logistic map and Henon map are shown in Table 1.

Table 1 shows that the number of 0 and the number of 1 of the two chaotic digital sequences are both approximately equal. But in terms of the long sequence, the degree of Henon sequence is smaller. That is to say, the digital Henon sequence can better satisfy the balance test requirement.

3.2 Run Test

Run test is usually used to determine if the number of run 1 or run 0 satisfies the requirements of the pseudo-randomness of digital sequences. The number of k -length run is about $\frac{1}{2^k}$ of the whole run in the same sequences. The run test results of digital sequences of Logistic map and Henon map are shown in Table 2. And each length of the sequences is 10^6 .

Table 2. Run test of digital sequences of Logistic map and Henon map

Run test	Logistic chaotic sequence	Henon chaotic sequence
number of 1-length run	246900	248765
Proportion of the total run	0.4962	0.4981
number of 2-length run	125642	125745
Proportion of the total run	0.2525	0.2518
number of 3-length run	62261	62625
Proportion of the total run	0.1251	0.1254
number of 4-length run	31469	31156
Proportion of the total run	0.0632	0.0624

From Table 2 we can find that 1-length run nearly satisfies the theoretical value $\frac{1}{2}$, 2-length run nearly satisfies the theoretical value $\frac{1}{2^2}$, 3-length run nearly satisfies the theoretical value $\frac{1}{2^3}$, 4-length run nearly satisfies the theoretical value $\frac{1}{2^4}$. But the values of Henon sequence are closer to theoretical values. So the digital Henon sequence can better satisfy the run test requirement.

3.3 Autocorrelation Test

The purpose of autocorrelation test is used to detect the correlation between sequences at a certain time and that at another time. The autocorrelation coefficient is defined as follows:

$$R(m) = \frac{1}{N} \sum_{i=1}^{N-m} x(i)x(i+m) \tag{4}$$

where N is the length of the sequence, and m is the value of the step length. The autocorrelation results of digital sequences of Logistic map and Henon map are shown in Fig. 9 and Fig. 10 respectively. And each length of the sequences is 10^6 .

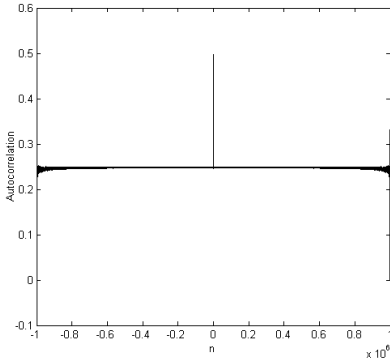


Fig. 9. Autocorrelation result of Logistic map

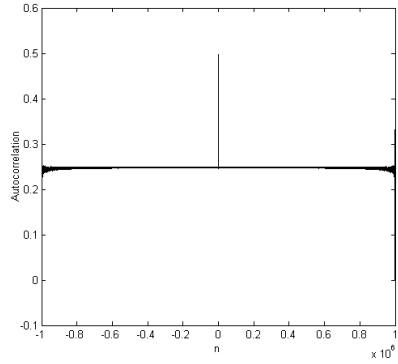


Fig. 10. Autocorrelation result of Henon map

As Fig. 9 and Fig. 10 show, the sharp peak is at 0, and at other time the correlation is small. That is to say, the period of the each sequence is great. Since the length is 10^6 , so each period is at least 10^6 . Thus, the digital sequences are similar to pseudo-random sequences.

4 Hardware Resources Utilization

In recent years, with the rapid development of digital circuits, the application of chaos theory to hardware equipments has gradually attracted people’s attention [20]. Now more popular devices include FPGA, DSP, etc., and because of the high logic density, versatility, low cost and other characteristics, FPGA is widely used in modern electronic technology. Furthermore, the models built in this paper are transformed into VHDL files. They can be used in FPGA. So the comparison of hardware resources is based on Cyclone II EP2C8Q208C8N. Firstly, transform the two models into VHDL files by using the SignalCompiler module.

Table 3. Hardware utilization of Logistic map

	Used	All	% of all
Total logic elements	246	8256	3%
Total combinational functions	246	8256	3%
Dedicated logic registers	37	8256	<1%
Total registers	37		
Total pins	4	138	3%
Total virtual pins	0		
Total memory bits	0	165888	0%
Embedded Multiplier 9-bit elements	16	36	44%
Total PLLs	0	2	0%

Table 4. Hardware utilization of Henon map

	Used	All	% of all
Total logic elements	812	8256	10%
Total combinational functions	804	8256	10%
Dedicated logic registers	111	8256	1%
Total registers	111		
Total pins	4	138	3%
Total virtual pins	0		
Total memory bits	0	165888	0%
Embedded Multiplier 9-bit elements	36	36	100%
Total PLLs	0	2	0%

Secondly, run the two projects in Quartus II environment, and then the resources that are used will be obtained. Table 3 and Table 4 are the hardware utilization of Logistic map and Henon map respectively.

If people want to apply chaotic sequences to hardware devices, the hardware resources will be an important aspect being worth considering. And the less the resources are used, the lower the cost consumes. From Table 3 and Table 4 we find that the resources Henon map makes use of are much more than Logistic map. Thus, in terms of conservation of resources, Logistic map is better.

5 Conclusion

This paper describes the whole process of building chaotic models and comparing the digital sequences at great length. According to chaotic equations, the models are built by the DSP Builder modules in Simulink library. And the digital chaotic sequences are obtained. Then these two sequences are tested and compared based on the statistical properties, including balance test, run test and autocorrelation test. Besides, the hardware utilizations are also compared. Eventually, we arrive at a conclusion that digital Henon sequences have better pseudo-randomness, but Logistic project uses less hardware resources. Thus, a chaotic map is chosen to apply depending on the different needs and may be taken full advantage of.

References

1. Zhan, M., Li, G.P.: Property Analysis of a Chaotic Sequence Generating Scheme and Its Improvement. *J. Southwest University (Natural Science Edition)* 30, 148–151 (2008)
2. Liu, B., Zhang, Y.Q., Liu, F.L.: A New Scheme on Perturbing Digital Chaotic Systems. *J. Computer Science* 32, 71–74 (2005)
3. Zhang, Y.P., Zuo, F., Zhai, Z.J.: A Color Image Encryption Algorithm Based on Chaotic Chebychev and Variable Parameters Logistic Systems. *J. Northwestern Polytechnical University* 28, 628–632 (2010)
4. Ding, L.N., Ding, Q., Chen, Q.: Design and Statistical Tests for Lorenz Chaotic Sequences Based on FPGA. *J. Electron Devices* 30, 1654–1657 (2007)
5. Jin, C.H., Yang, Y., Qi, C.D.: A Related-Key Attack on Chaotic Stream Ciphers. *J. Electronics & Information Technology* 28, 410–414 (2006)
6. Jin, C.H., Yang, Y.: A Divide-and-Conquer Attack on Self-synchronous Chaotic Ciphers. *J. Acta Electronica Sinica* 34, 1337–1341 (2006)
7. Zhang, F.: A New Six-Dimensional Chaotic Algorithms and Its Application in Image Encryption. *J. Microelectronics & Computer* 30, 62–65 (2013)
8. Wang, J.L., Cheng, L.Y.: Bifurcations and Chaos of the Model of Rossler System. *J. Henan Science* 30, 1403–1406 (2012)
9. Han, J.Q., Lun, S.X.: Method of Realizing Chaotic Masking Communication Based on Duffing System. *J. Computer Science* 40, 82–84 (2013)
10. Fan, J.L., Zhang, X.F.: Piecewise Logistic Chaotic Map and Its Performance Analysis. *J. Acta Electronica Sinica* 37, 720–725 (2009)
11. Tang, X.M., Zhao, D.F., Tan, M.C.: Study on an Improved Digital Chaotic Sequence of the Tent Map Applied in the DSSS. *J. Yunnan University (Natural Science Edition)* 32, 396–399 (2010)
12. Sun, H.Z., Mao, A.X., Su, X.Y., Liu, L., Wei, R.H., Liu, G.: Simulation of the Henon System's Dynamical Behavior's by Utilizing MATLAB. *J. Shangqiu Teachers College* 27, 54–57 (2011)
13. Zhao, X.Z., Li, Y.R.: Research on the Digital Image Encryption Technology Based on the Logistic Sequence. *J. Zhanjiang Normal College* 30, 88–91 (2009)
14. Zhang, H.L., Min, F.H., Wang, E.R.: The Comparison for Lyapunov Exponents Calculation Methods. *J. Nanjing Normal University (Engineering and Technology Edition)* 12, 5–9 (2012)
15. Liao, D.W., Zhu, W.Q.: Research on Lyapunov Exponents Algorithm and Its Application. *J. Wenzhou Vocational & Technical College* 8, 39–41 (2008)
16. Luo, L.J., Li, Y.S., Li, T., Dong, Q.T.: Research and Simulation of Lyapunov Exponents. *J. Computer Simulation* 22, 285–288 (2005)
17. He, S.L., Huang, Y., Huang, J.: Simulation Study on the Properties of Henon Mapping Sequences. *J. Kunming University* 34, 81–83 (2012)
18. Liu, X.K., Sun, X.H., Wan, L.H.: A New Image Encryption Based on Henon Mapping. *J. China Jiliang University* 19, 338–341 (2008)
19. Yu, Y.H., Liu, W.D.: Analysis of Balance of Chaotic Spreading Spectrum Sequences Based on Logistic-Map and Tent-Map. *J. Chongqing University of Posts and Telecommunications (Natural Science)* 16, 61–64 (2004)
20. Liao, N.H., Gao, J.F.: The Chaotic Spreading Sequences Generated by the Extended Chaotic Map and Its Performance Analysis. *J. Electronics & Information Technology* 28, 1255–1257 (2006)

Part V
**Intelligent Analysis for Biological,
Mobile and Cloud Computing**

Wavelet-Domain Image Watermarking Using Optimization-Based Mean Quantization

Huang-Nan Huang¹, Der-Fa Chen², Chiu-Chun Lin², and Shuo-Tsung Chen^{1,*}

¹Department of Applied Mathematics, Tunghai University, Taichung 40704, Taiwan (R.O.C.)
{nhuang, shough33}@thu.edu.tw

²Department of Industrial Education and Technology, National Changhua University of
Education, Changhua 500, Taiwan (R.O.C.)
dfchen@cc.ncue.edu.tw, anvy651018@yahoo.com.tw

Abstract. This study presents an image watermarking scheme that uses optimization-based mean quantization in the wavelet domain. In the proposed scheme, multi-coefficients of DWT are utilized for image watermarking. To modify these coefficients in mean quantization technique, an optimization formula is derived. First, the peak signal-to-noise ratio (PSNR) is expressed as a performance index in matrix form. Then, an optimized-quality functional that relates the performance index to the mean-quantization technique is obtained. Finally, the Lagrange Principle is utilized to obtain the optimal solution. The optimal solution is applied to watermarking. Experimental results show that the watermarked image can keep high PSNR and achieve better BER even when the number of coefficients for embedding a watermark bit increases.

Keywords: image, optimization-based, mean quantization, DWT, PSNR.

1 Introduction

In order to protect the copyright of an image, digital image watermarking that is based on the wavelet transform [1, 2] has become mature. In this study, DWT LH3 and HL3 coefficients are used to embed the watermark. Peak signal-to-noise ratio (PSNR) and bit error ratio (BER) are commonly performance indexes in measuring the quality and robustness of an image watermarking scheme. We use the proposed optimized-quality quantization watermarking scheme to balance the tradeoff between them [3-5]. First, the PSNR is rewritten as a performance index. An optimization functional is then proposed to relate this performance index to the mean-quantization technique. Finally, the Lagrange Principle is utilized to obtain the optimal solution. The optimal solution is applied to watermarking. In addition, the watermark can be extracted without the original image. The performance of the proposed scheme is evaluated by PSNR and BER. Experimental results show that the watermarked image using the proposed scheme can keep high PSNR and achieve better BER even when the number of coefficients for embedding a watermark bit increases.

* Corresponding author.

The rest of this paper is organized as follows. Section II reviews some preliminaries. Section III first rewrites PSNR as a performance index. An optimized-quality functional that relates the performance index to the mean-quantization technique is then proposed. Finally, the Lagrange Principle is used to solve the optimized-quality problem and the solution is applied to watermarking. Section IV does some experiments to test the performance of the proposed scheme. Conclusions are finally drawn in Section V.

2 Preliminaries

In this section, we review discrete wavelet transform for later use. The wavelet transform maps a function which belongs to functional space $L^2(R)$ onto a scale-space plane. The wavelets are obtained by a single prototype function $\psi(t)$ which is regulated with a scaling parameter and shift parameter [1, 2]. In any discretized wavelet transform, there are only a number of wavelet coefficients for each bounded rectangular region. Still, each coefficient requires the evaluation of an integral. To avoid this numerical complexity, one needs an auxiliary function, i.e., the basic scaling function $\varphi(t)$. The basic scaling function and the wavelet basis function are as follows.

$$\varphi_{j,n}(t) = 2^{j/2} \varphi(2^j t - n) \quad (1)$$

$$\psi_{j,n}(t) = 2^{j/2} \psi(2^j t - n) \quad (2)$$

where j and n are the dilation and translation parameters. A method to implement DWT is a filter bank that provides perfect reconstruction.

3 The Proposed Watermarking Scheme

In this section, the proposed optimization-based embedding and extraction using mean quantization technique is introduced.

3.1 Embedding Process

First of all, the watermark $B = \{\beta_i\}$ is randomly generated using a binary PN sequence. Hence the watermark values belong to the set $\{1,0\}$ and is adopted as the secret key K_1 . After the host image is transformed by applying DWT, k values of the DWT-coefficients are grouped into a column vector form:

$$C = [c_1 \ c_2 \ \cdots \ c_k]^T \quad (3)$$

The proposed embedding technique is described as follows. Suppose the amplitudes of the coefficients in C_j (say, the j^{th} group) are quantized to

$$z_j = \left\lfloor \frac{WC_j}{q} + \frac{1}{2} \right\rfloor \quad (4)$$

where $\lfloor \cdot \rfloor$ indicates the floor function, $q \in \mathbb{R}^+$ is the quantization size which is adopted as another secret key K_2 , and $W = \begin{bmatrix} 1 & 1 & \dots & 1 \\ k & k & \dots & k \end{bmatrix}$ is an $1 \times k$ matrix.

The proposed embedding rules are in the following:

- If $z_j \pmod{2} = \beta_j$, the amplitude of the coefficients in C_j is quantized to

$$y_j = \zeta_1 = z_j \times q \quad (5)$$

- If $z_j \pmod{2} \neq \beta_j$ and $z_j - \left\lfloor \frac{WC_j}{q} \right\rfloor = 0$, the the amplitude of the coefficients in C_j is quantized to

$$y_i = \zeta_2 = (z_j + 1) \times q \quad (6)$$

- If $z_j \pmod{2} \neq \beta_j$ and $z_j - \left\lfloor \frac{WC_j}{q} \right\rfloor \neq 0$, the the amplitude of the coefficients in C_j is quantized to

$$y_j = \zeta_3 = (z_j - 1) \times q \quad (7)$$

According to Eqs. (5) and (7), the embedding technique can be rewritten as an equation of the form:

$$g(\bar{C}_j) = WC_j - \zeta_1 = 0, \text{ if } z_j \pmod{2} = \beta_j, \quad (8)$$

or

$$g(\bar{C}_j) = W\bar{C}_j - \zeta_2 = 0, \text{ if } z_j \pmod{2} \neq \beta_j \text{ and } z_j - \left\lfloor \frac{WC_j}{q} \right\rfloor = 0 \quad (9)$$

or

$$g(\bar{C}_j) = W\bar{C}_j - \zeta_3 = 0, \text{ if } z_j \pmod{2} \neq \beta_j \text{ and } z_j - \left\lfloor \frac{WC_j}{q} \right\rfloor \neq 0 \quad (10)$$

where \bar{C}_j is the watermarked wavelet-coefficient vector that corresponds to C_j .

Let \mathbf{C} denote the collection of all groups of coefficients $C_j, j = 1, 2, \dots$, i.e., $\mathbf{C} = [C_1 \ C_2 \ \dots]^T$, and the corresponding weight matrix \mathbf{W} and \mathbf{Y} become

$\mathbf{Y} = [y_1 \ y_2 \ \dots]$ and $\mathbf{W} = [W_1 \ W_2 \ \dots]$, respectively. Then Eqns. (8), (9), and (10) can be grouped into

$$g(\bar{\mathbf{C}}) = \mathbf{W}\bar{\mathbf{C}} - \mathbf{Y} = 0$$

Generally, the quality of a watermarked image is evaluated by peak signal to noise ratio (PSNR) which is introduced as follows. If $\mathbf{I}(i, j)$ and $\bar{\mathbf{I}}(i, j)$ are the values of the original and corresponding modified pixels (i, j) in the original image \mathbf{I} and watermarked image $\bar{\mathbf{I}}$, respectively, then PSNR is defined as

$$PSNR = -10 \log_{10} \left(\frac{\sum_{i=1}^m \sum_{j=1}^n (\bar{\mathbf{I}}(i, j) - \mathbf{I}(i, j))^2}{255^2 mn} \right), \tag{11}$$

where m and n represent the height and width of the image.

Since the DWT is implemented with orthogonal wavelet bases, PSNR is expressed as

$$PSNR = -10 \log_{10} \left(\frac{\|\bar{\mathbf{C}} - \mathbf{C}\|_2^2}{255^2 mn} \right) \tag{12}$$

which can be rewritten as a performance index :

$$f(\bar{\mathbf{C}}) = \frac{\|\bar{\mathbf{C}} - \mathbf{C}\|_2^2}{255^2 mn} \tag{13}$$

or

$$f(\bar{\mathbf{C}}) = \frac{(\bar{\mathbf{C}} - \mathbf{C})^T (\bar{\mathbf{C}} - \mathbf{C})}{255^2 mn} \tag{14}$$

Based on the performance index $f(\bar{\mathbf{C}})$ in Eq. (14) and the constraint $g(\bar{\mathbf{C}})$ in Eq. (8), the optimization-based quantization problem has the following form:

$$\text{minimize } f(\bar{\mathbf{C}}) = \frac{(\bar{\mathbf{C}} - \mathbf{C})^T (\bar{\mathbf{C}} - \mathbf{C})}{255^2 mn} \tag{15a}$$

$$\text{subject to } g(\bar{\mathbf{C}}) = \mathbf{W}\bar{\mathbf{C}} - \mathbf{Y} = 0 \tag{15b}$$

To embed the watermark \mathbf{B} , we numerically solve the optimization problem (15) by using Matlab ToolBox.

3.2 Extraction Process

To detect the watermark, every k consecutive coefficients are grouped into $\bar{C}_i^* = \{|\bar{c}_1^*|, |\bar{c}_2^*|, \dots, |\bar{c}_k^*|\}$, where the superscript $*$ denotes the optimal result with respect to the corresponding variable. Then, the embedded binary bits are extracted by using the following rule.

$$\hat{\beta}_i = \left\lfloor W\bar{C}_i^* / q + \frac{1}{2} \right\rfloor \pmod{2} \quad (16)$$

Finally, the hidden watermarks are extracted as $\hat{B} = \{\hat{\beta}_i\}$ without the original audio signal. In other words, the proposed scheme is a blind watermarking scheme.

4 Experimental Results

This section presents experimental results that indicate the performance of the proposed image watermarking scheme. The host images, each of size 512×512 , are decomposed into three levels by applying DWT, and then the watermark is embedded into the LH3 and HL3 coefficients.

4.1 Image Quality Assessment and Embedding Capacity

To evaluate the quality of the watermarked image, four images, *Lena*, *Jet*, *Peppers*, *Cameraman*, are adopted as the example images. Figures 1-3 show the watermarked

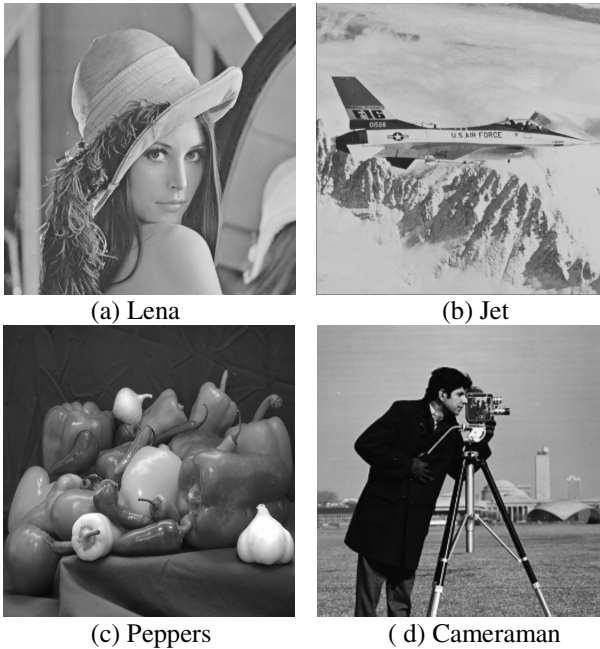


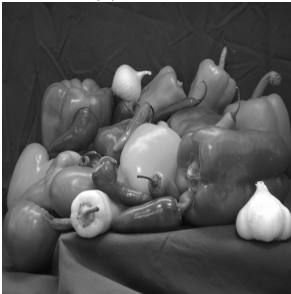
Fig. 1. Original images



(a) Lena



(b) Jet



(c) Peppers

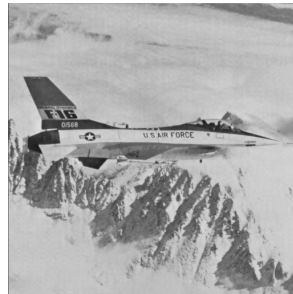


(d) Cameraman

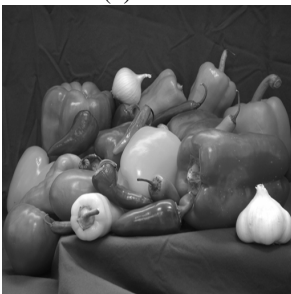
Fig. 2. Watermarked images for $k=2$, $q=26$



(a) Lena



(b) Jet



(c) Peppers



(d) Cameraman

Fig. 3. Watermarked images for $k=4$, $q=52$.

Table 1. PSNR and Embedding Capacity

Method	Image genre	Parameters	PSNR (dB)	Embedding capacity (bits)
Proposed method	Lena	$k=2, q=26$	42.5	4096
		$k=4, q=52$	42.4	2048
	Jet	$k=2, q=26$	42.6	4096
		$k=4, q=52$	42.3	2048
	Peppers	$k=2, q=26$	42.1	4096
		$k=4, q=52$	42.2	2048
	Cameraman	$k=2, q=26$	42.1	4096
		$k=4, q=52$	42.3	2048

images under various parameters. Table I shows the results of the image quality evaluation and the embedding capacity under various parameters.

4.2 Robustness Testing

To evaluate the robustness of the proposed method, 100 images including the four images, *Lena*, *Jet*, *Peppers*, *Cameraman*, are tested. After the embedding process, four attacks are adopted to test the robustness of the embedded watermark in case $k=2, q=26$ and $k=4, q=52$. The robustness is measured by Bit error ratio (BER) defined by

$$\text{BER} = \frac{B_{\text{error}}}{B_{\text{total}}} \times 100\%$$

where B_{error} and B_{total} denote the number of error bits and the number of total bits, respectively. Usually, an image is compressed before it is transmitted over the Internet. In addition, noise may come into the image during the transmission. The robustness test results against JPEG2000 compression and Gaussian noise are given Tables II and III, respectively.

Table 2. BER in JPEG2000 Compression

JPEG Quality factor		3	5	7	10	12
Proposed method BER	$k=2, q=26$	28.27	7.70	5.31	1.32	0
	$k=4, q=52$	15.11	3.71	3.02	0.76	0

Table 3. BER in Gaussian Noise

Method & parameter		dB		
		35	30	25
Proposed method BER	$k=2, q=26$	0	0.05	2.93
	$k=4, q=52$	0	0	0

5 Conclusion

This study presents an optimized-quality mean-quantization scheme for image watermarking. The watermark is embedded in the LH3 and HL3 coefficients of discrete wavelet transform (DWT). Based on an equation that connects PSNR and mean quantization technique, we obtained an optimization-based formula for image watermarking. Experimental results show that the watermarked image can keep high PSNR even the number of coefficients for embedding a watermark bit increases. In addition, the proposed scheme also achieves better BER.

References

1. Daubechies, I.: Ten Lectures on Wavelets. SIAM, Philadelphia (1992)
2. Burrus, C.S., Gopinath, R.A., Gao, H.: Introduction to Wavelet Theory and Its Application. Prentice-Hall, New Jersey (1998)
3. Chen, S.-T., Huang, H.-N., Chen, C.-J., Wu, G.-D.: Energy-proportion based scheme for audio watermarking. IET Proceedings on Signal Processing 4(5), 576–587 (2010)
4. Chen, S.-T., Wu, G.-D., Huang, H.-N.: Wavelet-domain audio watermarking scheme using optimization-based quantization. IET Proceedings on Signal Processing 4(6), 720–727 (2010)
5. Huang, H.-N., Chen, S.-T., Kung, W.-M., Lin, M.-S., Hsu, C.-Y.: Optimization-Based Embedding for Wavelet-Domain Audio Watermarking. Springer: Journal of Signal Processing Systems, doi:10.1007/s11265-013-0863-y

The Sybil Attack in Participatory Sensing: Detection and Analysis

Shih-Hao Chang¹, Kuo-Kun Tseng², and Shin-Ming Cheng³

¹ Department of Computer Science and Information Engineering,
Tamkang University, New Taipei City, Taiwan

`sh.chang@ieee.org`

² Department of Computer Science and Technology, Harbin Institute of Technology,
Shenzhen, China

`kchtseng@hitsz.edu.cn`

³ Department of Computer Science and Information Engineering,
National Taiwan University of Science and Technology,
Taipei City, Taiwan

`smcheng@mail.ntust.edu.tw`

Abstract. Participatory sensing is a revolutionary paradigm in which volunteers collect and share information from their local environment using mobile phones. Nevertheless, one of the most important issues and misgiving about participatory sensing applications is security. Different from other participatory sensing application challenges who consider user privacy and data trustworthiness, we consider network trustworthiness problem namely Sybil attacks in participatory sensing. Sybil attacks is a particularly harmful attack against participatory sensing application, where Sybil attacks focus on creating multiple online user identities called Sybil identities and try to achieve malicious results through these identities. In this paper, we proposed a Hybrid Trust Management (HTM) framework for detecting and analyze Sybil attacks in participatory sensing network. Our HTM was proposed for performing Sybil attack characteristic check and trustworthiness management system to verify coverage nodes in the participatory sensing. To verify the proposed framework, we are currently developing the proposed scheme on OMNeT++ network simulator in multiple scenarios to achieve Sybil identities detection in our simulation environment.

1 Introduction

In recent years, participatory sensing become a very popular and promising new technology to enable economically urban data sharing solution to a variety of application. Different from last century, the mobile phone of today, namely smartphone, have usually come with multiple embedded sensors, such as camera, microphone, GPS, accelerometer, digital compass and gyroscope. These technologies empowered smartphone users to collect data from their surrounding environment and upload them to an application server using existing communication infrastructure (e.g., 3G service or WiFi access points). Smartphones

provide an excellent platform for participatory sensing application [1]. Hence, a requester of data can create tasks that uses the general public to capture geo-tagged images, videos, or audio snippets. Participants who have installed the client APPs on their smart phones can submit their data and get rewarded. For example, Panoramic 3-D photosynthesis of businesses and restaurants photos from Gigwalk has been collected by Microsoft Bing Map.

Participatory sensing provides a very openness which allows anyone to contribute data, however, also exposes the applications to malicious and erroneous attack. Security in participatory sensing is complicated by the data sharing nature and the lack of tamper-resistant mechanism. In addition, due to the broadcast nature of the participatory sensing, it is impractical to rendering public key cryptography in distributed network environment. Sharing sensed data tagged with spatial-temporal information could reveal a lot of personal information, such as user's identity, personal activities, political views, health status, etc., which poses threats to the participating users. Hence, an attacker can have many identities to act maliciously, by either stealing information or provide incorrect data in participatory sensing environment, namely Sybil Attack. The Sybil attack was first introduced by Microsoft researcher J. R. Douceur [6]. A Sybil attack relies on the fact that a participatory sensing network data server cannot ensure that each unknown data collecting element is a distinct, mobile phone. Therefore, any malicious participatory sensing network attack can try to inject false information into the network to confuse or even collapse the network applications.

In cloud computing , everything is treated as a service (i.e. XaaS), e.g. SaaS (Software as a Service), PaaS (Platform as a Service) and IaaS (Infrastructure as a Service) and these services define a layered system structure for cloud computing. However, trust management is one of the most challenging issues in the emerging cloud computing. Although many approaches have been proposed recently for trust management in cloud environments, not much attention has been given to determining the credibility of trust feedbacks. To solve this problem, we propose a Hybrid Trust Management (HTM) framework for evaluating the trustworthiness of volunteer networks in participatory sensing applications. Our HTM framework allows a credit calculator associate with mobile devices that reflects the level of trust perceived over a period of time. A high credit score is an indication that a particular mobile device has been reporting reliable communication in the past. Moreover, in analyzing Sybil attacks data, we applied a fuzzy logic approach that can analyze detected Sybil attack features with these learning patterns in production time. To verify our idea, we utilizing OMNeT++ simulation to show its effectiveness against Sybil attacks.

The rest of this paper is organized as follows. Section 2 presents related works and summarized. Section 3 provides the detection factors to motivate the need for a reputation system in the context of participatory sensing and presents an overview of the system architecture respectively. In Section 4, we describe the experimental setup. Section 5 concludes the paper.

2 Background

In recent years, more and more participatory sensing applications apply in different fields. For example, in personal health monitoring, BALANCE [2] provides allows the client to monitor the activities and behavior of their diet, and encourage healthy living. It is the use of mobile phones enters the food calories and accelerator detects movement patterns and time to project the calories consumed to achieve health management. HealthSense [3] automatically detect health-related events, such as pain or depression cannot be observed directly through the current sensor technology. HealthSense analyze sensor data from the patient by machine learning techniques. The system uses patient input events to assist in classification (such as pain or itching). Finally, user provides feedback to the machine learning process. As mentioned, participatory sensing applications are exposes the applications to malicious an erroneous attacks.

The first Sybil attack was described by Douceur in the context of peer-to-peer networks [6]. He showed that there is no practical solution for this attack and pointed out that it could defeat the redundancy mechanisms of distributed storage systems. Problems arise when a reputation system (such as a trusted certification) is tricked into thinking that an attacking computer has a disproportionately large influence. Grover [7] proposed a scheme to protect against the Sybil attack using neighboring nodes information. In this approach every node participate to detect the suspect node in the network. Every mobile node have different group of neighbors at different time interval. After sharing their tables they match their neighboring table, if some nodes are simultaneously observed with same set of neighbors at different interval of time, then these node are under Sybil attack. In this case, identities are neighboring nodes that associated to specific trust devices. Similar to a central authority creating certificates, there are few ways to prevent an attacker from attaining multiple devices.

The concepts of trust and reputation have been shown to be promising concepts to support the customers in such situations in selecting a high quality service [8]. Trust and reputation are similar concepts and in computational models both are often based on history of past interactions. However, building up trust and reputation usually requires long-term identifiers which can be link over numerous transactions. At a first glance, this seems to be in conflict with the protection of the user's privacy, as unlinkability is a key term when referring to privacy properties. To solve this problem, Bayesian trust models [9] naturally allow for the interpretation of trust as a subjective probability, which allows for the consideration of personal preferences and context-dependent parameters. However, building up trust and reputation usually requires long-term identifiers which can be link over numerous transactions.

Due to in the participatory sensing applications, participants allowed anyone with an appropriate device that gets the application installed to a register as a participant. Such kind of human intervention entail serious security and privacy risks. Human behavior will involve additional security challenges. User's sensor data unboundedly transmit could results in leak of privacy. For instance, user may leak his/her personal identity information by nature of personal response.

Due to user may receive incorrect data from network that will lead integrity problem as it comes from malicious participants. For example, the malicious user can tamper and report data to other participants [4]. However, in the participatory sensing, introduces different security issues because devices are already in the hands of potential adversaries.

Trust management is one of the critical issues in cloud computing and a very active research area [10,11]. Brandic *et al.* Over the past few years, many studies have proposed different techniques to address trust management issues. For instance, [10] proposed a centralized approach in cloud environments using compliant management to help the cloud service consumers to support the cloud service consumer's perspective in selecting proper cloud services. Unlike previous works that use centralized architecture, they present a credibility model supporting distributed trust feedback assessment and storage. This credibility model also distinguishes between trustworthy and malicious trust feedback. Hwang *et al.* [11] proposed a security aware cloud architecture where trust negotiation and data coloring techniques are used to support the cloud service provider perspective. The cloud service consumer's perspective is supported using the trust-overlay networks to deploy a reputation-based trust management.

From the machine learning method aspect, the machine learning method is seek interesting failure patterns in the training data and then identify the failure symptoms with these learning patterns in production time. Fuzzy Knowledge Based Control [13] is an original interesting research area based on machine learning method. Fuzzy Knowledge Based Control objective is to modelling human reasoning in a simple method using a knowledge based system and an inference procedure. Because of Fuzzy Knowledge Based Control are based on fuzzy logic approach that depends on expert system for dynamic system control, large memory space and high speed computing capability are basic requirements.

In this section, we overview the state of art projects that design and the implementation of the participatory sensing, cloud computing framework and trust models. However, their approaches are not applicable to detect Sybil attacks in participatory sensing environments by utilize trust management system. Therefore, we attempt to identify Sybil attacks in participatory sensing environment by utilizing a Hybrid Trust Management (HTM) system that distinguish between credible trust nodes' feedbacks and malicious trust nodes' feedbacks through a credibility model.

3 Detection the Sybil Attack in Participatory Sensing Factors

We refer to the participants, i.e., smartphone users, in the system as entities. Interactions are actions between entities, i.e., the usage of a service or a capability that is offered by a service provider, e.g., buying goods or information. Hence, the type of interaction specifies the service context, in which a smartphone user, namely entity A, wants to interact with a service provider. Whenever, an entity A is in the role of the initiator of an interaction, i.e., entity A has to select a service

provider from a set of available service providers, it may evaluate the trustworthiness of the available service providers as a basis for the selection. Hereby, entity A uses its direct evidence from previous interactions and recommendations (also called indirect evidence). Having collected direct evidence and recommendations about one or multiple service providers, the trust model can be used for aggregating the evidence removing or giving lower weight to recommendations from unreliable sources and deriving trust values for the service providers, which then can be the basis for the decision whether to interact with one of the available service providers at all, and which service provider to select.

We propose a cloud based service management framework implemented in a service provider using the Service Oriented Architecture (SOA) to deliver trust as a service. SOA and Web services are one of the most important enabling technologies for cloud computing in the sense that resources (e.g., software, infrastructures, and platforms) are exposed in clouds as services. In particular, our framework uses Web services to interact with several distributed smartphone nodes that expose interfaces so that trust participants (other smartphone nodes) can give their trust feedbacks or inquire about the trust results based on feedback messages. Fig. 1, depicts the framework, which consists of three different layers, namely the Cloud Service Provider Layer, the Trust Management System Layer, and the Cloud Service Consumer Layer.

- The Cloud Service Provider Layer consists of different cloud service providers who provide cloud services. The minimum indicative feature that every cloud service provider should have is to provide the infrastructure as a service (i.e., the cloud provider should have a data center that provides the storage, the process, and the communication).
- The Trust Management System Layer. This layer consists of several distributed Trust Management System (TMS) nodes that expose interfaces so that cloud service consumers can give their trust feedbacks or inquire about the trust results represents.
- The Cloud Service Consumer Layer. Finally, this layer consists of different cloud service consumers who consume cloud services. For example, a new startup that has limited funding can consume cloud services (e.g., hosting their services in Amazon S3). A cloud service consumer can give trust feedbacks of a particular cloud service by invoking the TMS.

As mentioned, participatory sensing nodes are exposed to malicious participants may deliberately contribute forge nodes and bad data. These malicious participants also can exploit these links to de-anonymize the volunteers and compromise their privacy. Like other networks, the security requirements in participatory sensing include services such as authentication, confidentiality, integrity, and access control such as Sybil attack and slandering should be addressed. Once the Sybil attack participatory sensing, a Sybil node impersonating multiple identities has an important feature that can be detected by knowing the characteristics. For example, all the identities are part of the same physical device, they must move in unity way, while independent nodes are free to move

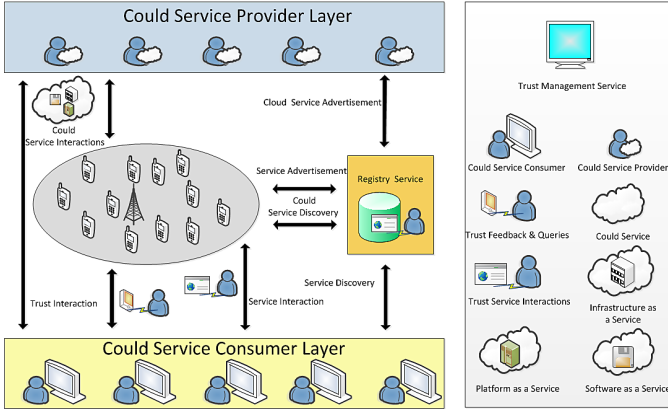


Fig. 1. Architecture of the Trust as a Service Framework

at will. As nodes move geographically, all the Sybil identities will appear or disappear simultaneously as the attacker moves in and out of range.

We develop this HTM to exploit Sybil attack characteristics to perform Sybil attack detection based on the following two assumptions:

- First, we assume that each user and service provider who wants to participate in the system owns a unique, initial identifier, which is obtained at the bootstrapping phase from a party that is trusted by all involved parties (i.e., users, service directory provider, and service providers).
- Second, we assume the Sybil nodes uses a single-channel radio, multiple Sybil nodes must transmit serially, whereas multiple independent nodes can transmit in parallel.

Characteristics Checking Scheme. This HTM framework include a passive Characteristics Checking Schemes (CCS) that keep Sybil nodes in check including time, density and topology in the simultaneously. The idea of this CCS introduces an adaptive threshold (similar as the watchdog implementation method) to detect the characteristics of a Sybil attacks in participatory sensing network. This CCS will be implemented in the cloud-side which regularly check the coverage participatory sensing nodes condition to decide whether the node either genuine identity or has been compromised. The CCS will set multiple adaptive thresholds to monitor covered participatory sensing nodes' characteristics and implemented as part of the system operation process running on the cloud server. When a requester inquire trust credit to an inspector from HTM framework, if the passive CCS does not detect any attack pattern on the node, it returns no attack pattern found to requester. Otherwise, it will notify requester to disconnect suspicious malicious node(s).

Time. Once a Sybil node has compromising a partial participatory sensing, it will create a number of online identities and use these identities to compromise

participate sensing. Hence, in utilizing server statistics number of connected participants for a brief period of time, we can first distinguish between suspect Sybil attacks in participatory sensing network. By analyzing this statistics, we can infer system whether has suspicious Sybil nodes at that time period. We assume current number of connected participatory is S_c , statistics number of connected participatory at this time is S_r , and set threshold ϵ . Detective method is defined as follows.

$$\frac{S_c}{S_r} = \begin{cases} \text{It could has some dubitable node,} & \text{if } \frac{S_c}{S_r} > \epsilon, \\ \text{It could has no any dubitable node,} & \text{if } \frac{S_c}{S_r} \leq \epsilon. \end{cases} \quad (1)$$

As shown in Fig. 2, we assume this system’s threshold ϵ is 2 with S_c and S_r are 100 and 40 at T8. As presented in (1), we can know while S_c divided by S_r is greater than ϵ . In this situation, the system can assume the suspected Sybil nodes existed in participatory sensing network.

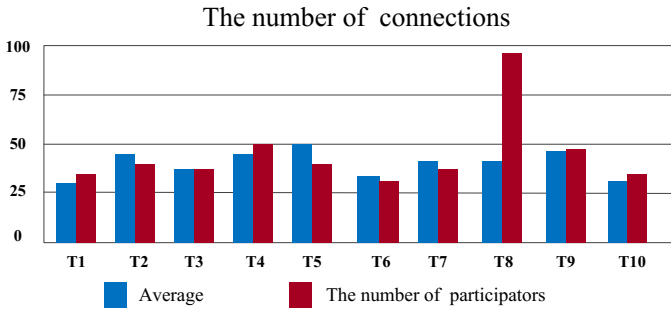


Fig. 2. A diagram of suspicious Sybil attacks activities for a short period of time

Density. Moreover, after filtered the time factor to monitor suspected Sybil identities, our passive detection scheme will based on the fundamental assumption that the probability of two mobile users having exactly the same set of neighbors in a sub-region and its topographical map will smaller than 1000m x 1000m [12]. Each sub-region usually has regular density, hence we can exploit this characteristic to detect suspected Sybil nodes. Using server statistic each region’s density for a brief period of time. We assume each sub-region is inside a base station coverage range. By this statistics report, we can infer system whether has suspected Sybil node in his sub-region. we assume current region’s density is D_c , statistics region’s density is D_r , and set threshold θ . Detective method is defined as follows.

$$\frac{D_c}{D_r} = \begin{cases} \text{It could has some dubitable node,} & \text{if } \frac{D_c}{D_r} > \theta, \\ \text{It could has no any dubitable node,} & \text{if } \frac{D_c}{D_r} \leq \theta. \end{cases} \quad (2)$$

As shown in Fig. 3, left is statistics sub-region’s density and right is current sub-region’s density. Nodes that has a mark “S” are a suspected Sybil identities,

so we can observe current sub-region's density is greater than statistics sub-region's density in a brief period of time. In this situation, the system can assume the suspected Sybil nodes existed in participatory sensing network.

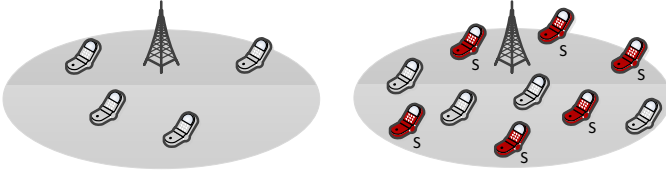


Fig. 3. A diagram of suspicious Sybil attacks activities in a region

Network Topology. Due to each Sybil group will present a similar topography map, nodes will be very frequently heard together even when they are not Sybil identities and will rarely be heard apart as they do not move out of radio range. This leads to the false identification rate in topographies that are denser in terms of nodes per square meter. Hence, the accuracy and error rates for a single node observer when a Sybil attacker present will be very obvious. Again, in smaller topographies there is insufficient mixing to separate Sybil identities from real nodes, and the error rate is high, as is the detection rate, because all nodes are seen as part of the same identity. As the topography size increases, the number of meaningful observations that a single node can make increases, and the true positive rate stays high, on the order of 95%, while the false positive rate drops significantly. As the topography size increases further, the number of observations that a single node can make is reduced as all nodes are spread far apart, and the accuracy of identifying the Sybil identities decreases.

As shown in Fig. 4, when Sybil attacks is present, the network topology can be conceptually divided into two parts: one consisting of all genuine identities and the other consisting of all Sybil identities. The link connecting a genuine node to a Sybil node is called an attack edge [13].

Trust Credit Assessment. In our framework, the trust credit of a participatory sensing node is evaluated by our Trust Credit Assessment (TCA) scheme. Its represented by a collection of officiate history records denoted as H . Each requester node r holds her point of view regarding the trustworthiness of a inspector node i in the officiate history record which is managed by a trust management service. Each officiate history record is represented in a tuple that consists of the participatory sensing node primary identity P , the inspector node identity I , a set of trust credit T and the aggregated trust feedbacks weighted by the credibility Tc (i.e., $H = (P, I, T, Tc)$). Each credit in T is represented in numerical form with the range of $[0, 1]$, where 0, +1, and 0.5 means negative feedback, positive feedback, and neutral respectively.

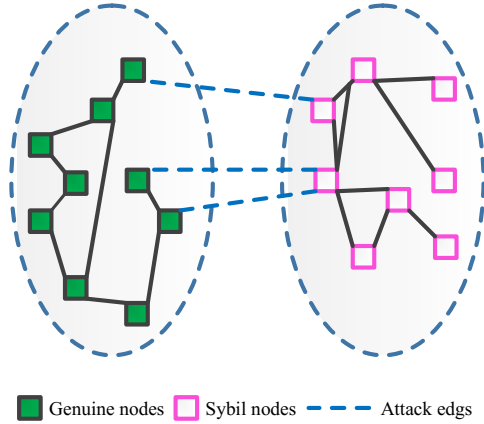


Fig. 4. A network topology diagram of suspicious Sybil attacks activities

Whenever a requester node inquires the trust management service regarding the trustworthiness of a inspector node i , the trust result, denoted as $Tr(i)$, is calculated as the following:

$$Tr(i) = |v(i)|Tc(l, i)/|v(i)|, \tag{3}$$

where $V(i)$ is all of the feedbacks given to the inspector node i and $|V(i)|$ represents the length of the $V(i)$ (i.e., the total number of feedbacks given to the inspector node i). $Fc(l, i)$ are the trust feedbacks from the l th cloud consumer weighted by the credibility.

Analytical Decision Making. The idea of our analytical decision making introduces the notation of a thread-based checkpoint to detect the abnormal behavior of a participatory sensing node communication. Rather than addressing the misbehaviors of these participatory sensing nodes, Sybil identity checker is to analysis and identify the compromised participatory sensing nodes based on the adaptive threshold. This Sybil identity checker can be embedded as a part of the communication process and linking list running on the participatory sensing node to monitor participatory sensing node communication behavior. Similar as the watchdog implementation method, Sybil identity checker monitor the participatory sensing node behavior by implementing the threads function running on the sensor nodes. By registering a checkpoint, the defining thread is expected to set that checkpoint with the declared timeout value. Each thread registers one or more reputation checker to be expected, stored in either a communication process or linked list. When checkpoint has been registered, it sets the observer points Vobs (the set of observable) inside the system process and communication links. To achieve identify the Sybil node behavior in the sensor system; the Sybil identity checker applies the adaptive threshold to verify the

measure result. This adaptive threshold use J_0 as the measurement symptoms computed under normal operating conditions (U_0, X_0).

$$J_0 = J_0(U_0, X_0), \tag{4}$$

And $J(U, X)$ the measurement thresholds used for verify the Vobs measurement results. We suggest threshold according to:

$$J(U, X) = J_0(U_0, X_0) + \Delta J(U - U_0, X - X_0), \tag{5}$$

where ΔJ represent increase or decrease according to the operating conditions and the range will be decided by the program designer. $J(U, X)$ the measurement threshold are then generated when comparing the Vobs measurement results with the updated thresholds. Once the Vobs measurement results is not inside the range of the measurement threshold $J(U, X)$ ($Vobs \supset J(U, X)$), the fault can be identified by the Sybil identity checker scheme. Due to the faults may occur at different levels in the wireless sensor networks, in this paper, we only concern node-level Sybil identity identification during sensor system operation. Once the Sybil identity checker identified the behavior of the Sybil identity, it will notify others participatory sensing nodes to execute segment of Sybil identities from participatory sensing network.

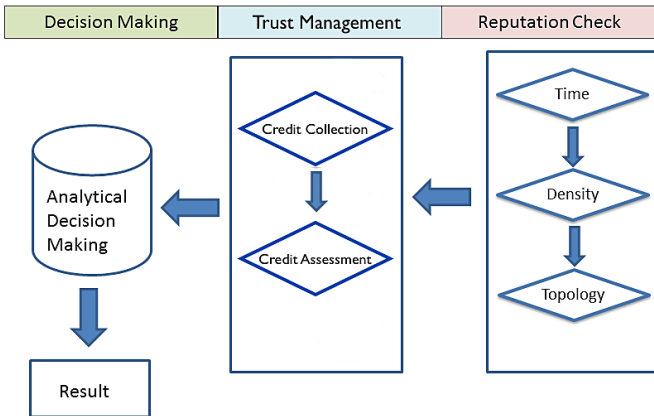


Fig. 5. Hybrid reputation monitoring diagram

An Example of Scenario. As this attack has no relation to the identification scheme, we do not further evaluate it. On the other hand, an attacker can be compromised and controlled by a Sybil node. This compromised genuine node is considered as a Sybil node and not as a genuine node. This Sybil node will focus on creating multiple online user identities called Sybil identities and try to achieve malicious results through these identities. As shown in Fig. 5, we will proceed in three phases. In the first phase, the server-side manager defined multiple adaptive thresholds including time, density and network topology to evaluate

network trustworthiness. When multiple Sybil identities has been identified operation exceed adaptive threshold range in our CCS, CCS module will generate a notification to the TCA. Then TCA will officiate this inspector node history records from its database and process the credit assessment. Once the Sybil attacks pattern has been preliminary identified, it will enable Analytical Decision Making (ADM) to further analysis and determine the Sybil attacks in this network. This framework will check regularly network and system's statistics and use adaptive threshold to achieve network trustworthiness.

4 Experimental Evaluations

In this section, the proposed algorithm Hybrid Trust Management (HTM) scheme will be present and implement in OMNeT++ [14]. OMNeT++ is an extensible, modular, component-based, C++ simulation library and framework which also includes an integrated development and a graphical runtime. It provides a generic component architecture based on object oriented approach. Model components are termed modules which primarily communicate with each other via message passing either directly, or via pre-defined conditions and the message can arrive from another module or from the same module. We are currently implementing the CCS modules on each mobile participants as it well depicts a real world situation. This mobility model is based on entity mobility model where the nodes move independent of each other. We have taken following parameters for implementation as shown in Table 1.

Table 1. Simulation Parameter Setup

Parameter	Values / Ranges
Simulation area	5000m × 5000m
Simulation time	1000s
Speed (m/s)	0.0 m/s to 5.0 m/s
Routing protocol	GPRS
Number of nodes (Max)	1000
Number of base stations	1-3
Traffic source	CBR
Pause time	Uniformly distributed in 0-50s
Packet size	256 bytes
Packet rate	5 packets/s
Transmission range	3000m

5 Conclusion

In this paper, we proposed a Hybrid Trust Management (HTM) framework for detecting Sybil attacks in participatory sensing network. Our HTM was proposed for performing trust management and reputation checker to verify coverage nodes in the participatory sensing. Sybil attacks focus on creating multiple

online user identities called Sybil identities and try to compromise system with its malicious results through these identities. This HTM framework combined two schemes namely Characteristics Checking Scheme (CCS) and Trust Credit Assessment (TCA) to a suspicious Sybil node observation. CCS was proposed for passively monitor suspected Sybil nodes characteristics including time, density and topology in the participatory sensing network simultaneously. TCA was proposed for evaluate trustworthiness of the suspected Sybil nodes. We are currently working on actual system testing to evaluate network performance in the detect of Sybil nodes based on OMNeT++.

References

1. Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., Srivastava, M.B.: Participatory sensing. In: Proc. ACM WSW (October 2006)
2. Denning, T., Andrew, A., Chaudhri, R., Hartung, C., Lester, J., Borriello, G., Duncan, G.: BALANCE: Towards a usable pervasive wellness application with accurate activity inference. In: Proc. ACM HotMobile 2009 (February 2009)
3. Stuntebeck, E.P., Davis II, J.S., Abowd, G.D., Blount, M.: HealthSense: Classification of health-related sensor data through user-assisted machine learning. In: Proc. ACM HotMobile 2008 (February 2008)
4. Deng, L., Cox, L.P.: LiveCompare: grocery bargain hunting through participatory sensing. In: Proc. ACM HotMobile 2009 (February 2009)
5. Mendez, D., Labrador, M.A.: On sensor data verification for participatory sensing systems. *Journal of Networks* 8(3), 576–587 (2013)
6. Douceur, J.R.: The Sybil attack. In: Druschel, P., Kaashoek, M.F., Rowstron, A. (eds.) IPTPS 2002. LNCS, vol. 2429, pp. 251–260. Springer, Heidelberg (2002)
7. Grover, J., Gaur, M.S., Laxmi, V.: A Sybil attack detection approach using neighboring vehicles in VANET. In: Proc. SIN 2011, pp. 151–158 (November 2011)
8. Josang, A., Ismail, R.: The Beta reputation system. In: Proc. 15th Bled Electron. Commerce Conf. (June 2002)
9. Ries, S.: Extending Bayesian trust models regarding context-dependence and user friendly representation. In: Proc. ACM SAC 2009, pp. 213–237 (March 2009)
10. Brandic, I., Dustdar, S., Anstett, T., Schumm, D., Leymann, F., Konrad, R.: Compliant Cloud Computing (C3): Architecture and language support for user-driven compliance management in clouds. In: Proc. IEEE CLOUD 2010 (July 2010)
11. Hwang, K., Li, D.: Trusted cloud computing with secure resources and data coloring. *IEEE Internet Computing* 14(5), 14–22 (2010)
12. Piro, C., Shields, C., Levine, B.N.: Detecting the Sybil attack in ad hoc networks. In: SecureComm 2006 (August 2006)
13. Chang, S.-H., Huang, T.-S.: A fuzzy knowledge based fault tolerance algorithm in wireless sensor networks. In: Proc. IEEE AINA 2012, pp. 891–896 (March 2012)
14. Hornig, R., Varga, A.: An overview of the OMNeT++ simulation environment. In: Proc. SIMUTools 2008 (2008)

Vessel Freeboard Calculation Method Based on Laser Scanning

Yingce Zhao, Guangming Lu, Xiaotang Guo, and Yazhuo Wang

Shenzhen Graduate School, Harbin Institute of Technology, China
luguangm@hit.edu.cn

Abstract. In the inland river shipping business, vessel freeboard value is considered as the primary standard to measure whether the ship is overloaded or not, so vessel freeboard calculation methods become an important topic to concern. This paper presents a vessel freeboard measurement scheme based on laser scanning technology, which can calculate the freeboard value with high accuracy by combining with K-MEANS clustering analysis and Hough transform. This method has been applied to the actual inland river shipping monitoring system and has gotten good performance.

Keywords: Vessel freeboard, Laser sensor, K-MEANS, Hough transform.

1 Introduction

In the inland river shipping industry, vessels often try to load more goods to earn more money. But this kind of situation may cause severely incident, such as running aground, wreck, etc. Until now, researchers have presented different methods to monitor whether the vessel is overload or not, but they still have some limitations. These methods are listed as follows:

(1) Pressure sensor method: Take advantage of the feature that changing of water pressure can reflect the changing of water depth, the pressure sensor is mounted at the waterline of vessel. When the vessel is loaded, the output of pressure sensor will be changed, and then the freeboard can be calculated by some mapping functions [1].

(2) Ultrasonic method: Ultrasonic has been widely used in people's daily life, such as speed measurement, distance measurement, etc. The principle by adopting ultrasonic technology to detect the tonnage of the vessel is mainly to fuse the data of speed sensor and ultrasonic sensor, thus to evaluate the tonnage of the vessel. The ultrasonic sensor array are installed at both sides of the river to detect the vessel cross section and calculate its area, then the interval data scanned by ultrasonic transducer and the vessel speed data obtained by speed sensor will be fused, thus the vessel sailing distance can be calculated, and then the vessel underwater volume can be gotten.

(3) Image processing method: In this method, the water gauge of vessel is captured by imaging device, then we can use image processing technology to determine whether the vessel is overloaded or not by identifying the vessel's water gauge readings.

(4) Multiple visual information fusion technology: inland waterways monitoring system track vessel localizer and speed when the vessel enters surveillance. After determining the vessel's position, one or more images of vessel are captured by the imaging device. At the same time, the vessel image is classified by the system, and then monitoring data may be gotten by image processing method [2].

(5) Multi sensor data fusion technology: multi infrared sensors are installed at both sides of the channel in detecting area, and the sonar array are installed on riverbed wall, and at the detecting area side and top is equipped with two camera devices, and a signal processing circuit which is used for receiving the signal from infrared sensor and controlling sonar scanning, and after receiving the sonar signal, the signal processing circuit will transfer it to computer. At the same time, the captured vessel image data is also sent to computer. After fusing all data mentioned above, the actual load tonnage and total tonnage can be got in real-time [3].

In the method of pressure sensor and image processing, the additional devices should be installed on the vessel, which make them work in poor conditions and easily to be damaged. Besides, the vessel owners may change or destroy the devices to avoid supervision. Another disadvantage of image processing method is that the accuracy is very low since it will be affected by weather and light conditions. When the sonar technology is used to measure the freeboard of vessel, the devices should be installed on the riverbed, and will be easily covered by silt and be difficult to maintain. In conclusion, the above methods have different weak points, and we need find new method to solve these problem.

In the last decade, laser sensor technology has been got a huge development, and has been widely used in distance measurement [4,5] and speed detection[6,7]. In this paper, we present a novel vessel freeboard measurement scheme based on laser scanning technology, which can calculate the freeboard value with high accuracy by combining with K-MEANS clustering analysis and Hough transform methods.

2 Framework and Methodology

In the shipping industry, the different types of vessel have different rated freeboard. We can determine whether the vessel is overload by comparing the actual freeboard with rated freeboard. Vessel rated freeboard values can be obtained directly from the vessel specification. In this paper, we propose an actual freeboard measurement scheme by using laser sensors, which is indicated in Fig.1.

In Fig. 1: h_1 is water height, h_2 is deck height, h_3 is the monitoring value of freeboard, L is a straight line distance between the laser sensor and the measured vessel, r is the deviation angle of L from the vertical direction. Freeboard value can be calculated as equation (1):

$$h_3 = h_1 - h_2 \quad (1)$$

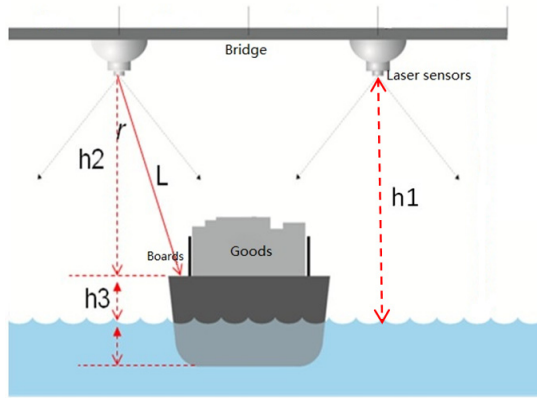


Fig. 1. Freeboard measurement scheme

2.1 Data Preprocessing

When the vessel goes through under the laser sensors, the sensors start periodical scanning, and the minimal angle difference of two scanning point is 0.25° , and the scanning angle range is from -5° to 185° , then we can get 761 sampling points in one cycle. Each sampling points is corresponding to the distance from scanned point on cross-section of the vessel to the sensor. Usually, it is difficult to find the significant characteristic of the deck in one single data cycle, so multiple period scanning is required to obtain a complete vessel scanned data. In this paper, 1100 cycles of data is taken as a unit for freeboard measurement.

The data returned from laser sensor is the straight line distance L from laser sensor to the object. We should convert it into the vertical distance. Sensor coordinate system is shown in Fig. 2.

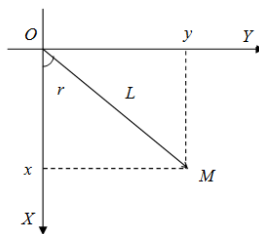


Fig. 2. Sensor coordinate system

M is the measured point on vessel cross-section or on water surface, which is expressed as (L, r) in the laser sensor coordinate system. Desired vertical distance can be obtained by equation (2):

$$|ox| = L \times \cos r \tag{2}$$

The further optimization of data processing is needed after the distance transformation. Bijection relationship is established between vertical distance data and binary image pixel, which is expressed as $Sensor_{ij} \leftrightarrow Binary_{ij}$, where $i=1, 2 \dots 1100$ means sensor's acquisition cycle, $j=1, 2 \dots 761$ represents the number of sampling points per cycle. When the sensor data is 0, binary data is processed to 0, otherwise 1.

To improve the accuracy and speed of processing and analyzing of the vessel's deck data, the water data from $Sensor_{ij}$ should be removed. The difference between water data and deck data can be gotten through the observation and analysis of the scanned data. Because of the waves, the measured water data values are floating up and down, and there is no same value in one cycle, while the measured deck data value is stable because the deck is smooth. Use such difference, we can remove water data. Sliding window technology is used to process $Sensor_{ij}$, and the detailed steps are as follows.

- 1) According to sensor measuring characteristics, the measured data associated with the deck covers the distance of 3-6 points, so set the window size as 5.
- 2) Take five non-zero points of each cycle in sampling sequence to store in the sliding window.
- 3) Analysis on the data of sliding window: If the floating is relatively large, then treat the data as the water data, remove the first point, go to 4); If the floating is relatively small, then treat the data as deck data, do not need do any treatment;
- 4) Take the first non-zero point in the subsequent analysis into the sliding window, fill sliding window, and go to 3).

Usually, the complicated weather condition, such as fog, rain and dust in the air, may cause noise in the $Binary_{ij}$, so the noise filtering step is needed, here we use the median filter to remove the noise. The comparison of the binary image before and after data preprocessing is shown in Fig. 3.



Fig. 3. Comparison of the $Binary_{ij}$ before (left) and after (right) preprocessing

2.2 Vessel Freeboard Calculation

Calculate Water Height h_1 . After the data preprocessing, $Sensor_{ij}$ is gotten. We can easily find the vessel data, and then the remaining data is water data. Water height h_1 can be got by following steps:

- 1) $Sensor_{ij}$ is vertical distance which is obtained in preprocessing data module. The initial set of weights for each data point is $P_{ij}=1$.
- 2) Calculating the average value V of valid data:

$$V = \frac{1}{\sum_{i=1}^N \sum_{j=1}^M P_{ij}} \sum_{i=1}^N \sum_{j=1}^M Sensor_{ij} P_{ij}, \quad i = 1 \dots 60, j = 1 \dots 761 \quad (3)$$

- 3) Update the weights P_{ij} by removing discrete points. Set the $Sensor_{ij}$ in the range of threshold $(V - h, V + h)$ as valid data, set $P_{ij}=1$, otherwise $P_{ij}=0$. Recalculated

$$V = \frac{1}{\sum_{i=1}^N \sum_{j=1}^M P_{ij}} \sum_{i=1}^N \sum_{j=1}^M Sensor_{ij} P_{ij} \quad (4)$$

- 4) Repeat 2), 3) until V is stable, V is the desired water height h_1 .

Calculate Deck Height h_2 . Since decks locate on both sides of the vessels and occupy relatively regular fields, we can use Hough transform to find the edges of one vessel, then the decks can be located near the edges. Usually, the decks are 100cm wide, and there will be 3-6 laser points during one cycle, and the heights of all deck points are almost the same, the clustering method is adopted to find all the points in the multiple cycles. Finally the height of deck h_2 can be gotten by calculating the mean height of the clustered points. The detailed steps are as follows:

Hough transform to extract deck. Sometimes, the returned data maybe not complete, since it may be affected by some external factors (weather, noise, etc.). Then the calibration step is needed. Here we use interpolation processing to calibrate the sensor data. The interpolation processing includes the following steps:

- 1) Finding a point $P_{(n,m)}$ satisfies $Sensor_{n,m} = 0$.
- 2) Calculating adjacent point V_1 and V_2 :

$$Y_1 \in \{y \mid sensor_{n,y} \neq 0, y < m\} \quad Y_2 \in \{y \mid sensor_{n,y} \neq 0, y > m\}$$

$$V_1 = sensor_{n,y} \quad y = \max(Y_1) \quad (5)$$

$$V_2 = sensor_{n,y} \quad y = \min(Y_2) \quad (6)$$

- 3)
$$P_{(n,m)} = \frac{V_1 + V_2}{2} \quad (7)$$
- 4) Repeat 1), 2), 3) until the scanning is finished.

The comparison of the vessel image before and after interpolation is shown in Fig. 4(a) and 4(b), respectively.

Then we use Canny operator to detect the deck edges of a vessel. Canny operator first select a certain Gaussian filter for image smoothing to eliminate noise, then refine the smoothed image gradient amplitude matrix to look for the possible edge points in the image, finally, find the image edge points through by dual-threshold detection method to complete the edge extraction. The image after Canny edge detection is shown in Fig. 4(c).

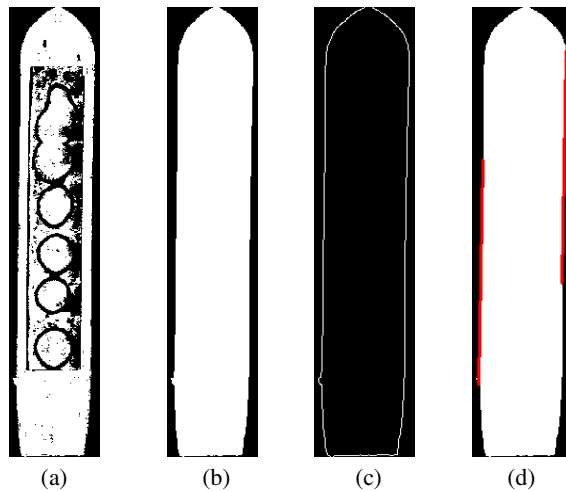


Fig. 4. (a) The original image of the vessel; (b) The image of the vessel after Interpolation processing; (c) The edge image obtained by canny operator; (d) Deck location by Hough transform

At last, we try to locate the deck by Hough transform. The Hough transform result is shown in Fig. 4(d), in which red line is used to indicate the outermost straight line found by Hough transform. Take the straight line at left side in Fig. 4(d) for example, extracting the values of the adjacent pixels at its two sides and the mapping $Sensor_{ij}$ to constitute the deck data distribution diagram, which is shown in Fig. 5, in which red points indicate the height value of red straight line, black and blue points indicate the height value left and right of the red line separately.

K-MEANS clustering algorithm to calculate the value of the freeboard. Sometimes the water data is not completely removed in preprocessing step, and also the sensor may return data with noise, or the points may be reflected from other part of the vessel. They will give us much difficult to find the deck points. So the points around the

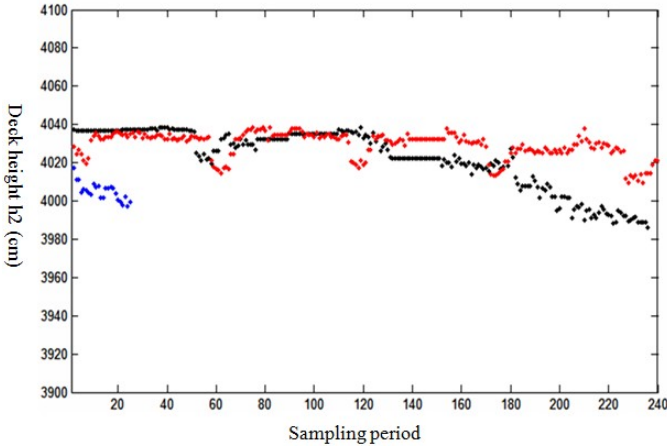


Fig. 5. Deck points obtained by Hough transform

edge generally include three kinds of data points: water data, deck data, and other parts of vessel. The relationship of these three kinds of points is that the value of water height is greater than the value of deck, and the value of deck height is greater than other parts' height. The height in each group is relatively close, and the height in different group is relatively different. According to these features, we can use clustering method to group the deck points out.

From Fig. 5, we can see that most of the deck points are concentrated around the red points, but there are also some interference points with deviations. Thus, the deck points can be found by K-MEANS clustering method, where the K is set as 3, the formula for calculating K-MEANS distance is shown as below:

$$d = |s_j - x_i|, j = 1 \dots 3, i = 1 \dots n \tag{8}$$

The cluster center is set to s_j , and extracting the values of vessel is set to x_i . The value of cluster center is deck height h_2 , the formula for calculating h_2 is shown as below:

$$h_2 = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i, j = 1 \dots 3 \tag{9}$$

$|C_j|$ represents the number of point in C_j . by equation (1), the value of the measured freeboard h_3 can be gotten.

3 Experiments

In order to prove the efficiency of the proposed method, we set up a testing database, in which 97 vessels' freeboards are measured by manpower and our automatic monitoring system, respectively. Under the request of the shipping administration section, the difference should be within ± 15 cm.

Part of the comparison results is shown in Table 1. From the table we can see that the automatic measure results are very close to the manual measure results.

The difference distribution of the 97 vessels is shown in Fig 6. 99% of the measuring differences are within $\pm 15\text{cm}$, and 84% is within $\pm 10\text{cm}$. It shows that the proposed method can meet the requirement of the real application.

Besides, the proposed method can run in real time, and can work in complicated weather conditions, such as raining, snowing, day and night. The system has been installed on a bridge of Changjiang River, China, and has worked for one year. It can monitor all the vessels when they go through the bridge, and give a warning when an overloaded vessel is detected. A camera is installed on the same bridge, and can capture the overloaded vessel at the same time, and then a record will be added into the database for further inquiry.

Table 1. Freeboard measurement results

Record	Manual measurement (cm)	Automatic measurement (cm)	Difference (cm)
Vessel 1	32	31.2317	0.7683
Vessel 2	63	51.9196	11.0804
Vessel 3	64	53.4376	10.5624
Vessel 4	443	445.683	-2.683
Vessel 5	8	2.21882	5.78118
Vessel 6	223	220.185	2.815
Vessel 7	79	81.315	-2.315
Vessel 8	142	141.168	0.832
Vessel 9	92	101.247	-9.247
Vessel 10	112	120.868	-8.868
Vessel 11	80	62.9053	-17.0947
Vessel 12	30	36.9367	-6.9367

Freeboard measurement results

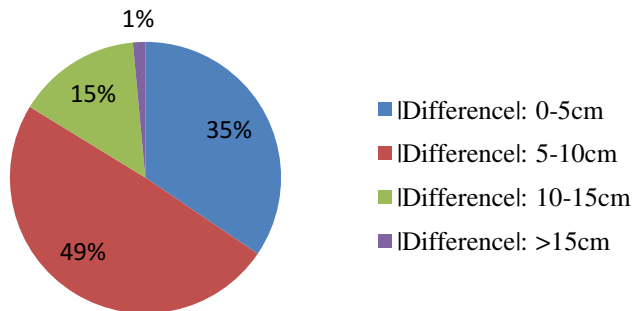


Fig. 6. The difference distribution of the 97 vessels

4 Conclusions

This paper propose a novel freeboard measuring scheme based on laser scanning, in which we use Hough transform method to find the deck, and use K-MEANS methods to cluster the deck points. Then the freeboard can be gotten by calculating the difference between the water high and the deck high. Comparing with current measurement methods, our method can run in real time, and has high accuracy, high efficiency and low cost.

Acknowledgments. The work is supported by the NSFC fund (61020106004, 61271344), and Key Laboratory of Network Oriented Intelligent Computation, Shenzhen, China.

References

1. Sun, G., Mao, Q.: Study on Automatic Determining Ship's Draft and Stability Parameters. *Navig. China* 2, 28–30 (2002) (in Chinese)
2. Meow, D.C.K.: Method and system for surveillance of vessels. United States Patent: 20060244826A1
3. Fan, X.: Intelligent measuring system and measuring method of tonnage of ship. European Patent: CN20061011508
4. Zhao, H., Shibasaki, R.: A Novel System for Tracking Pedestrians Using Multiple Single-Row Laser-Range Scanners. *IEEE Trans. Sys., Man Cybern.* 35, 283–291 (2005)
5. Cui, J., Zha, H., Zhao, H., et al.: Laser-based Detection and Tracking of Multiple People in Crowds. *Comput. Vis. Image Underst.* 106, 300–312 (2007)
6. Musayev, E.: Laser-based large detection area speed measure methods and systems. *Opt. Lasers Eng.* 45, 1049–1054 (2007)
7. Lee, K.H., Ehsani, R., Castle, W.S.: A laser scanning system for estimating wind velocity reduction through tree windbreaks. *Comput. Electron. Agric.* 73, 1–6 (2010)

Visual Information Analysis for Big-Data Using Multi-core Technologies

Nikolaos Mpountouropoulos, Anastasios Tefas, Nikos Nikolaidis,
and Ioannis Pitas

Artificial Intelligence and Information Analysis Laboratory,
Department of Informatics, 54124 Thessaloniki, Greece
{tefas,nikolaid,pitas}@aiia.csd.auth.gr
<http://www.aiia.csd.auth.gr/>

Abstract. The exponential growth of video data produced by surveillance cameras, cell phones and movie post-production creates the need to process big-data using methods that are able to produce instantaneous result. Video summarization can be accomplished and represented in several manners. The achieved summaries might be a sequence of images or short videos. In our method, an input video is divided into segments. From each segment we calculate key frames using three different key frame definitions, to summarize the video data. The contribution of this paper is to describe how to incorporate techniques that extract *on the fly* results.

Keywords: video summarization, big-data video analysis, mutual information.

1 Introduction

Videos are structured according to a descending hierarchy of scenes, shots, frames. Frame is the fundamental unit of a video. A shot is a consecutive sequence of frames captured by a static camera or a moving one. A scene is a sequence of shots. A scene is defined as a collection of one or multiple shots focusing in an object or objects that motivate our interest. Video summarization is the process of detecting the most important and informative frames of a video in order to create a shorter version of it that is still able to convey the original message. The representative frames are called key frames. The importance of video summarization has become really apparent in now days, due to the exponential growth of video production and consumption over the internet and security applications. Only at youtube.com 100 hours of video are uploaded every minute.

Video segments are separated at the point where a shot-cut point is detected. Shot-cut corresponds to an abrupt or gradual frame change. Shot-cuts are classified in two major categories [1]. In the first category belong the cases where transitions between the frames are abrupt and the second one includes the cases of

gradual transitions, such as fade in/fade out, dissolve,wipe etc [2]. Video segmentation is easier to detect in an abrupt change rather than in a gradual transition. In the case of dissolve the frames of the shot in a video start to fade out while the next appears and grows clearer as the first one dims. The wipe happens when one shot replaces another in a different spatial regions of the intermediate video frames. The former frames grow using a pattern ,e.g. like a star, of a special shape until it entirely replaces the latter. A fade in/fade out is a gradual disappearance of a frame into black and then the black frame fades into the appearing shot.

Generally video segmentation algorithms work by extracting features from frames, then using a similarity measure to detect them. Features used for video segmentation include color histogram [5],[6], block color histogram, motions vectors, etc. To measure the similarity between frames using the extracted features is the second step. The similarity metrics can be the Euclidean distance, the chi-squared similarity , mutual information, etc.

We studied many standard techniques for detecting video segments and key-frames. We decided to adopt MI (Mutual Information) for video segmentation as mentioned in [3]. Using MI values we determine the segments. Key frames are extracted from these segments using three different key frame definitions. In Section 2 we briefly describe the adopted video segmentation algorithm. In Section 3, we describe the key frame selection process. Experimental results are provided in Section 4. Conclusions are drawn in Section 5.

2 MI: Mutual Information

2.1 Definitions and Background

In probability theory the mutual information of two discrete random variables is a measure that evaluates the mutual dependence of the two. Let X be a discrete random variable with a set of outcomes $A_X = \{a_1, a_2, \dots, a_N\}$ with the corresponding probabilities $\{p_1, p_2, \dots, p_N\}$. $p_x(x = a_i) = p_i, p_i \geq 0$ and $\sum_{x \in A_x} p_X(x) = 1$. Entropy of X measures “unpredictability” and can defined as:

$$H(X) = - \sum_{x \in A_X} p_X(x) \log p_X(x) \quad (1)$$

The *joint entropy* of two discrete random variables X, Y can be obtained by:

$$H(X, Y) = - \sum_{x, y \in A_X, A_Y} p_{XY}(x, y) \log p_{XY}(x, y) \quad (2)$$

where $\log p_{XY}(x, y)$ is the joint probability density function for the random variables X, Y . The *conditional entropy* of Y given X (or X given Y accordingly) is expressed as:

$$H(Y|X) = \sum_{x \in A_X} p_X(x) H(Y|X = x) = - \sum_{x, y \in A_X, A_Y} p_{XY}(x, y) \log p_{XY}(x|y) \quad (3)$$

The *mutual information* (MI) of the two variables X and Y is defined by:

$$I(X, Y) = - \sum_{x, y \in A_X, A_Y} p_{XY}(x, y) \log \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)} \quad (4)$$

According to [3] $I(X, Y)$ in (4) is equivalent to

$$I(X, Y) = H(X) - H(X|Y) \quad (5)$$

In our experiments, we have adopted the MI definition in (5).

2.2 Video Segmentation Using Mutual Information

In our approach the MI is calculated separately for each one of the RGB components. We normalize each RGB pixel luminosity to gray level from $0, \dots, N - 1$. At frame f_t three vectors with dimension N created containing the values on gray level for each pixel. Dividing each element by total number of the pixels of the frame and averaging them gives $p_X(x)$ where $A_X = \{0, 1, \dots, N - 1\}$ contains corresponding probability of each gray-scale pixel. Thus we can calculate the entropy $H(X)$ using (1).

Between two frames f_t, f_{t+1} three $N \times N$ matrices, $C_{t,t+1}^R, C_{t,t+1}^G$ and $C_{t,t+1}^B$ created containing information on the gray-level transitions between these frames. Each one of the components for example $C_{t,t+1}^B(i, j)$ with $0 \leq i \leq N - 1$ and $0 \leq j \leq N - 1$ corresponds to the occurrence that a pixel with gray level i in frame f_t has gray level j in the frame f_{t+1} . Dividing by the total number of the pixels of the video frame we find the joint probability in each matrix $C_{t,t+1}^{JPR}, C_{t,t+1}^{JPG}$, and $C_{t,t+1}^{JPB}$, that a pixel with gray level i has now gray level j . For two frames f_t, f_{t+1} the matrix of gray scale joint probability is given by:

$$C_{t,t+1}^{JP} = (C_{t,t+1}^{JPR} + C_{t,t+1}^{JPG} + C_{t,t+1}^{JPB})/3 \quad (6)$$

Using the expression in (6) we can calculate the conditional entropy mentioned at (3). Finally we use the expression (5) to calculate the $I(X, Y)$ for the two frames.

In order to detect video segments we define a temporal window W size of N_W . Local MI mean values, and the standard deviation of them excluding the current value $I_{t,t+1}$ at the current window t_c is described at [8]:

$$\bar{I}_{t_c} = \frac{1}{N_W} \sum_{t \in W, t \neq t_c} I_{t,t+1} \quad (7)$$

If the quantity $\bar{I}_{t_c} - I_{t,t+1}$ is greater than the double of the standard deviation of the selected values a video segment is detected.

2.3 Fast Implementation of the Algorithm

In order to reduce the processing time required for this task, we have devised a parallel (multi-threaded) implementation of the original algorithm. We decided

to automatically split the video sequences in blocks (non-overlapping subvideos) and process each block independently in different threads. Furthermore, we have designed a parallel video reading/decompressing implementation. The adoption of this operation, further improved the computational speed of the algorithm.

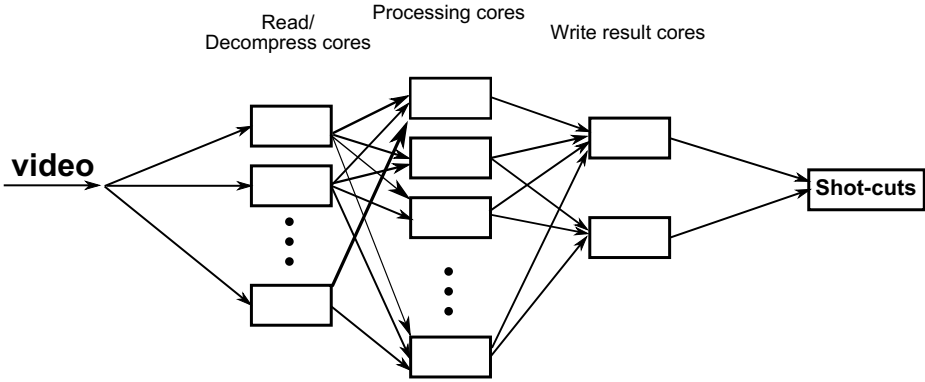


Fig. 1. Processing Cores

The first step for developing this algorithm was to evolve the *OpenCV* library. We created a new multi-threading method which can read and decompress the frames from the compressed video file using all the cores of the current working station. This way we can define the buffer of the processing frames per core. For example a buffer value equal to 60 means that each core will decompress 60 frames and write them to memory. Each one of these video-blocks is assigned to a thread responsible to calculate the MI from one frame to another. We define a matrix with size equal to the length of the movie. When the threads complete the calculations, each thread writes to the corresponding position of the matrix, the results of the corresponding calculations. Combing all these results using an adaptive thresholding approach we can define the segments of the video.

3 Key Frame Selection

The data entry for this algorithm can be composed of the video segments defined from the previous method. For each segment a key frame will be selected. Key frames are intended to be informative regarding the corresponding segment. According to this, a key frame is defined as the one that provides the smallest distance from the remaining frames in the segment. Three algorithm implementations are available. These are: *distance between frames*, *distance from the average frame* and *distance between frame histograms*.

All three of them need to compute distances between frames. For the two first algorithms, the distance between two frames is the sum of all their corresponding (having same coordinates) pixel distances. Pixel distances can be computed by

two methods using the distance of the *average* or the *euclidean* (9) pixel distance. The *average* distance can be defined, where $p_{R_t}, p_{G_t}, p_{B_t}$ is the pixel color RGB values of the current frame at position (x, y) and $p_{R_{t+1}}, p_{G_{t+1}}, p_{B_{t+1}}$ the next frame f_{t+1} , at the same position as:

$$d_{avg} = (|p_{R_t} - p_{R_{t+1}}| + |p_{G_t} - p_{G_{t+1}}| + |p_{B_t} - p_{B_{t+1}}|)/3 \tag{8}$$

The *average* pixel distance is the distance of their average values based on the RGB values of the pixel. The *euclidean* distance of two pixels is given by (9). The third algorithm calculates the distance between the histograms using correlation, chi-square, intersection and Bhattacharyya distance. Details on these metrics can be found in [4].

$$d_{euc} = \sqrt{(p_{R_t} - p_{R_{t+1}})^2 + (p_{G_t} - p_{G_{t+1}})^2 + (p_{B_t} - p_{B_{t+1}})^2} \tag{9}$$

3.1 Simple Distance of Frames

The former algorithm initially computes the distance for each segment frame pair (that is for frame pairs $1 - 2, 1 - 3, \dots, 2 - 1, 2 - 3, \dots$) where the distance between two frames is defined as the sum of their corresponding (having same coordinates) pixel distances, as mentioned above. Let $x_i \in \mathbb{R}^N, i = 1, \dots, M$ be the video frames in vectorized form. Let s_j be the key frame for the j -th segment. According to this we have:

$$s_k = \arg \min_i \sum_j \|x_i - x_j\|^2 \tag{10}$$

After all distances among segment frames are computed, the key frame can be derived as the one that has the smallest sum of frame distances, meaning that is the one closest to most other segment frames. The initial implementation of the algorithm was reading-decompressing two frames from the segment calculates their distance and stores it appropriately. Thus the complexity of reading-decompressing and comparison was $O(n^2)$. We make two significant improvements. We create a multi-thread version of the algorithm where each segment is being process by one core. Since the segments are unequal, when a thread finishes a new segment is being reassign to it. The reading-decompressing of each segment is done exactly only one time assuming that the working station has enough memory to fit in the segments being processed.

3.2 Average Frame

The second algorithm is computationally faster ($O(n)$ instead of $O(n^2)$ of the first method) but less accurate. We can average the frames of the segment to calculate the average frame of the segemnt by computing the average value of all corresponding frame pixel luminosities. The distance from the average frame is calculated by the Euclidean distance or the average distance as seen:

$$s_k = \arg \min_i \|x_i - x_{avg}\|^2 \tag{11}$$

where x_{avg} is the "avegare" frame.

After that, each frame is compared with the average frame of the segment, by using any type of distance mentioned before. In this case, key frames could be the ones that are the most similar (least distance) to the average ones. This is by far the fastest algorithm of the two but slightly less accurate. The complexity is $O(n)$.

3.3 Frame Histogram Distance

This algorithm follows a similar process to first to produce its key frames, with the only difference being that the distance between two frames in this algorithm is not the sum of their corresponding pixel distances, but the distance of their histograms in RGB color space. This algorithm is the most context sensitive of the three, and can yield drastically different results from the first two. The third algorithm we use is distance computation in a histogram level rather than pixel level, which is more robust to noise and camera movement. The distance function we use is correlation, chi-square, intersection and Bhattacharyya.

4 Experimental Results

In this section we describe experiments conducted in order to illustrate the decrease of execution time observed by our parallel implementation. We set the buffer values to 20, 40, 60 frames and change the number of threads for the MI algorithm as seen at Tables 2,3,4. The experiments were conducted on an Intel(R) Core(TM) i7-4770 CPU 3.40GHz PC which has 4 physical cores and can manipulate 8 logical threads. Each core processes a block of frames and writes the result to the corresponding position of a matrix. When the threads finish, the matrix contains the MI values for each pair of frames. We used Hollywood movie *Movie1* for the experiments of MI as seen at Table 1. Using four physical cores the ideal reduction of time will be 75%. According to [7] using the Hyper-Threading Technology we gain a little more time archiving at 79,72% setting the buffer to 60. The rest videos used as input were other Hollywood movies as seen at Table 1. The metrics for the distance algorithms *simple frame distance*, *average frame* was the average distance of the pixels, while in *histogram distance frame* was correlation.

Table 1. Characteristic of movies

	Total no.Frames	Resolution	Duration
Movie1	181764	960x540	2h 6m 21s
Movie2	196224	960x540	2h 16m 24s
Movie3	150361	960x540	1h 44m 31s

The algorithms of key frame selection especially the *simple distance frame* and *Histogram Distance Frame* are slower to our standards and takes more than the real time of the movie to complete. As seen at Table 5 we managed to

Table 2. Experimental results(buf=20) for MI

Number of Threads	1	2	4	6	8
Elapsed Time	34m 55s	20m 09s	12m 33s	10m 43s	9m 48sec
Decrease in time (%)		42,3%	64,1%	69,4%	72,1%

Table 3. Experimental results(buf=40) for MI

Number of Threads	1	2	4	6	8
Elapsed Time	33m 09s	15m 36s	9m 36s	8m 27s	7m 28sec
Decrease in time (%)		53%	71,1%	74,6%	77,5%

achieve an enormous reduction. Among the three of them the best time is the *Average Frame*. Although is not as robust as the other two the execution time is magnitude times faster than the other two as seen at Table(5). With *MI* and *Average Frame* can produce results on the fly.

Table 4. Experimental results(buf=60) for MI

Number of Threads	1	2	4	6	8
Elapsed Time	33m 41s	13m 35s	8m 29s	7m 38s	6m 50sec
Decrease in time (%)		59,68%	74,82%	77,74%	79,72%

Table 5. Time needed for key-frame selection in Movie1

	Simple Distance Frames	Average frame	Histogram Distance Frame
Single-core	1d 2h 52min 32sec	1h 3m 8sec	2d 1h 49min 33s
Multi-core	4h 59m 25s	32m 36s	9h 7m 11s

Table 6. Video segments and fastest method of key-frames

	Video segment using (MI)	Key frame selection using Average frame	Total Time
Movie1	6m 50s	32m 36s	39m 26s
Movie2	9m 33s	35m 13s	44m 46s
Movie3	6m 7s	27m 7s	33m 14s

We try to combine techniques for processing big-data video to accomplish instantaneously execution time. The fastest method for the selection of key frames, *average frame* might not be as robust as the others but completes calculation magnitude times faster. The next step is to examine algorithms to use those key frames to create a video summarization. The ideal is to adopt the proper algorithms for the creation of summary so that the total time of processing doesn't overrun the length of the movie.

Acknowledgment. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 316564 (IMPART). This publication reflects only the authors views. The European Union is not liable for any use that may be made of the information contained therein.

References

1. Hu, W., Xie, N.: A survey on visual content based video indexing and retrieval. *IEEE Transactions on Systems, Man, and Cybernetics* 41(6), 797–819 (2011)
2. Cotsaces, C., Nikolaidis, N., Pitas, I.: Video shot boundary detection and condensed representation: A review. *IEEE Signal Processing Magazine* 23(2), 28–37 (2006)
3. Cernekova, Z., Pitas, I., Nikou, C.: Information theory-based shot cut/fade detection and video summarization. *IEEE Transactions on Circuits and Systems for Video Technology* 16 (January 2006)
4. Opencv metrics for histograms, <http://docs.opencv.org/modules/imgproc/doc/histograms.html?highlight=comparehist#comparehist>
5. Smoliar, S.W., Zhang, H.J., Kankanhalli, A.: Automatic partitioning of full-motion video. *ACM Multimedia Syst.* 1(1), 10–28 (1993)
6. Cernekova, Z., Kotropoulos, C., Pitas, I.: Video shot segmentation using singular value decomposition. *SPIE Journal of Electronic Imaging* 16(4) (December 2007)
7. Chen, Y.K., Holliman, M., Debes, E., Zheltov, S., Knyazev, A., Bratanov, S., ... Santos, I.: Media Applications on Hyper-Threading Technology. *Journal Intel Technology* 6(1) (2002)
8. Pitas, I., Venetsanopoulos, A.: *Nonlinear Digital Filters: Principles and Applications*. Kluwer Academic (1990)

Application of Job Shop Based on Immune Genetic Algorithm

Lei Meng and Chuansheng Zhou

Software College, Shenyang Normal University, Shen Yang, 110142, China
netmenglei@126.com, 252752602@qq.com

Abstract. Job Shop scheduling problem, as an important part of computer integrated manufacturing system engineering, is a classic NP-hard combinatorial optimization problem and has vital effect on production management and control system. In this paper, base on biological immune system's antigen recognition, maintaining the diversity of antibodies and other features, a proposed improved genetic algorithm-the immune genetic algorithm is put forward, the algorithm will introduce the thinking of biological systems immune to the genetic algorithm, namely in use of first immune knowledge it structures inspection operator. By vaccination and immune selection, it not only retains the best individual groups but also ensures the diversity of individuals, thus avoiding the premature convergence of evolutionary search and improving convergence speed, meantime, an improved immune genetic algorithm, and adopting timely dynamic vaccination and the shut down criteria are given. Simulation results show that the algorithm is effective.

Keywords: Genetic Algorithm, immune, job shop, NP-hard, hypermutation.

1 Introduction

The job shop scheduling problem (JSP), may be described as follows: given n jobs, each composed of several operations that must be processed on m machines. Each operation uses one of the m machines for a fixed duration. Each machine can process at most one operation at a time and once an operation initiates processing on a given machine it must complete processing on that machine without interruption.

The operations of a given job have to be processed in a given order. The problem consists in finding a schedule of the operations on the machines, taking into account the precedence constraints that minimize the make span (C_{max}), that is, the finish time of the last operation completed in the schedule[1].

2 Genetic Algorithm Optimization Strategy

The foregoing model can be used find optimal values of several model parameters (design variables). Obviously an object function must be defined which determines the

quality of a certain set of design values. In genetic optimization the object function is often called the fitness function. Here the focus is on finding the distribution of stress near holes. Genetic algorithms are the best choice to solve such problems.

“A Genetic Algorithm (GA) is a programming technique that mimics biological evolution as a problem-solving strategy.” It is based on Darwinian’s principle of evolution and survival of fittest to optimize a population of candidate solutions towards a predefined fitness. [2]

GA uses an evolution and natural selection that uses a chromosome-like data structure and evolve the chromosomes using selection, recombination, and mutation operators. The process usually begins with randomly generated population of chromosomes, which represent all possible solution of a problem that are considered candidate solutions. Different positions of each chromosome are encoded as bits, characters or numbers. These positions could be referred to as genes. An evaluation function is used to calculate the goodness of each chromosome according to the desired solution; this function is known as “Fitness Function”. During evaluation, two basic operators, crossover and mutation, are used to simulate the natural reproduction and mutation of species. The Selection of chromosomes for survival and combination is biased towards the fittest chromosomes. [2][3]

A GA generally has four components. A population of individuals represents a possible solution. A fitness function which is an evaluation function by which we can tell if an individual is a good solution or not. A selection function decides how to pick good individuals from the current population for creating the next generation. Genetic operators such as crossover and mutation which explore new regions of search space while keeping some of the current information at the same time [4].

The following is a typical GA procedure:

```

Begin
  Initialize population;
  Evaluate population members;
  While termination condition not satisfied do
  Begin
    Select parents from current population;
    Apply genetic operators to selected parents;
    Evaluate offspring;
    Set offspring equal to current population;
  End
End

```

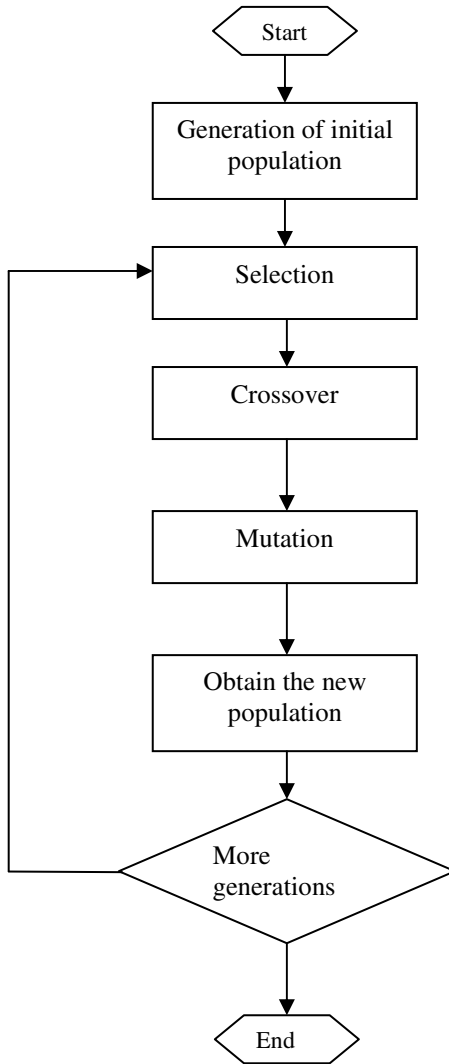


Fig. 1. Scheme of classical genetic algorithm

3 Immune System

The Biological information processing system can be divided into: brain systems, genetic and immune systems. People in practice the three major information systems through the simulation of biological research has been based on the three information processing system of the three intelligent algorithms that simulate brain systems based on artificial neural networks, genetic system based on simulation and simulation-based

genetic algorithm the immune system, artificial immune algorithm. In these three algorithms, artificial neural networks and genetic algorithms have been extensively studied so far, and was applied to various fields. However, the immune system structure based on artificial immune algorithm, due to the complexity of the immune system, the method difficult to design more complex, so the other two algorithms did not get equal attention, so at home and abroad and the application of research results is relatively small .Meanwhile, the three intelligent algorithms, genetic algorithms and artificial immune algorithm and relatively similar in both their genetic algorithm, mutation, crossover, belonged to the evolutionary algorithm, similarity is relatively large. Both have their own advantages and disadvantages. Such as the genetic algorithm is faster to maintain their individual diversity through crossover and mutation operators to achieve, but proved such a diversity of populations evolving with significantly reduced, the convergence of the algorithm easy to fall into local minimum during optimal solution and local search-based crossover operation ability, which is the genetic algorithm needs to be overcome. The use of artificial immune algorithm crossover and mutation operators to maintain the diversity of individual solutions, and the solution by increasing the concentration of antibodies and inhibition of regulatory mechanisms to promote the production of antibody solution, thus increasing the diversity of solutions of antibody selection, to further ensure the algorithm can converge to the global optimal solution, but the calculation is relatively slow. [5]

Artificial immune algorithm began in the 20th century, late 90s, now the research has just started. The idea comes from the biological immune system, which the immune system by simulating the learning, memory and other functions for pattern recognition and search optimization. Artificial immune algorithm will correspond to the antigen and antibody optimization objective function and feasible solutions. The affinity of the antibodies and antigens as a feasible solution to the matching degree with the objective function: the affinity between antibodies to ensure the diversity of feasible solutions, by calculating the expected survival rate of antibodies to promote optimum antibody of heredity and variation, with the memory cell unit After saving merit similar to the feasible solutions to curb continue to produce feasible solutions and accelerate the search to the global optimal solution, the same time, when similar problems recur, can quickly adapt to produce the optimum solution of the problem or even optimal solution.[5]

4 Algorithm Design

SGA prematurely and low search efficiency in the latter part of the problem, many improvements have been proposed operator method, in which the immune genetic algorithm is proposed based on biological immune mechanism An improved genetic algorithm that the actual objective function corresponds to solving the problem for the antigen, and antibodies corresponding to the solution of the problem. Biological immune system to antigens invading the body through the life of cell division and differentiation, automatically produce antibodies to protect against, a process known as immune response. In the immune response, some antibody preserved as memory cells invaded again when the same antigen, memory cells are activated and rapidly

produce a large number of antibodies, so again, faster response more strongly than the initial response, reflecting the memory function. Also, antibody mutual promotion between antibodies and inhibition, in order to maintain the diversity of antibodies and immune balance, this balance is based on the concentration of the mechanism, that is, the higher the concentration of antibodies, the more restrained; lower the concentration the more affected by promotion, reflects the self-regulation. On the basis of the existing genetic algorithm, the paper said the introduction of information entropy between the affinity and antibody concentration, so that it can more effectively express affinity between antibody and concentration, and using a new evaluation indicators - polymer affinity, can effectively ensure the diversity of antibody group, On the basis of the existing genetic algorithm, the paper said the introduction of information entropy between the affinity and antibody concentration, so that it can more effectively express affinity between antibody and concentration, and using a new evaluation indicators - polymer affinity, can effectively ensure the diversity of antibody group.[6]

The basic algorithm steps described below

- Step 1: Initialize the population, population size N set
- Step 2: If the primary response, the function within the parameters of the range of all randomly generated antibodies.
- Step3: For each antibody, the algorithm to calculate the fitness value.
- Step4: Genetic algorithms
- Step5: The antibody affinity group H , and set affinity antibody population threshold H_0 if $H < H_0$, Into the next generation of optimization operations, into Step 3; if $H > H_0$, P is a randomly generated antibodies, antibody at this time scale is $N + P$, to maintain the antibodies diversity.
- Step6: Determine whether the termination condition, if not, the next generation of group optimization, into Step 3; stable or if they meet the concentration of antibodies to the conditions at the end of algebra, the output the optimal solution.

5 Conclusion

Shop scheduling problem is that many experts now the focus of attention, genetic algorithm is used to solve the optimization problem of job shop scheduling algorithm, this paper, genetic algorithm optimization to improve the efficiency of in-depth and meticulous research, an improved immune genetic algorithm and the algorithm is applied to the workshop to solve scheduling problems.

References

1. Goncalves, J.F.: European Journal of Operational Research, 77–95 (2005)
2. Li, W.: Proceedings of the United States Department of Energy Cyber Security Group 2004 Training Conference, pp. 24–27 (2004)

3. Alhazzaa, L.: King Saud University Computer Science Collage CSC590_Selected Topic (2002)
4. Stein, G.: ACM Southeast Regional Conference Proceedings of the 43rd Annual Southeast Regional Conference, vol. 2, pp. 136–141 (2005)
5. Liu, X.Y.: Master's thesis, Project Management, Tianjin University (2008)
6. Wang, A.T.: Master's thesis, Communication and Information System, Ocean University of China (2008)

Study of Evaluation of GPS/BeiDou Combination Regional Navigation Satellite System

Tenghong Liu^{1,2} and Songlin Liu^{1*}

¹ Zhongnan University of Economics and Law,
School of Information and Safety Engineering,
182 Nanhu Avenue, Wuhan 430073, China

² Wuhan Yangtze Business University, Wuhan, China
liutenghong@21cn.com,
roger0007@sina.com

Abstract. 2020 with more than 30 satellites, and will provide global coverage. The paper compares Beidou open service signal B1 with conventional GPS signal L1. B1 has better performance in terms of multipath error and cross correlation noise. The paper uses the Beidou/GPS observations of the Beidou Experimental Tracking Stations to realize the Beidou and Beidou/GPS static and kinematic precise point positioning (PPP). The results will show that the combined system allows for improved integer ambiguity success rates, satellite visibility and reliability as compared to the systems separately.

Keywords: BeiDou, GPS, RTK, precision, variance matrix, reliability, Minimal Detectable Bias, integer ambiguity success rates.

1 Introduction

The BeiDou-2 satellites system navigation system is a Chinese self-developed regional active three-dimensional satellite positioning and communication system (CNSS), is in addition to the U.S. GPS, Russia's GLONASS third after mature satellite navigation system. CNSS effectively improve the accuracy of satellite positioning, navigation and timing performance capabilities. Beidou satellite navigation system, part of the three parts of the space segment, ground segment and user. Currently, the Beidou system consists of five fully operational geostationary satellite orbit (GEO) satellites in geosynchronous orbit and tilt 5 (IGSO) on three frequency bands (B1, B2, B3) transmit navigation signals in quadrature phase shift keying (QPSK) modulation. Based on a large number of scholars have carried out research PNT Compass [1-4], Shicheng et al [5] use of GNSS receivers (UB240-CORS) collected a large amount of real data on the quality and relative positioning measurements for analysis. Montenbruck et al [6] using his country produced tracks GNSS receiver for post-processing of data and clock products real data collected in the preliminary assessment.

* Corresponding author.

RTK (Real - time kinematic) real-time dynamic difference method is commonly used to locate a new measurement method, previously static, rapid static and dynamic measurements are needed after the solver to get centimeter-level accuracy, and is capable of RTK field measurements in real time to get centimeter-level positioning accuracy, which uses a dynamic real-time carrier phase differential method is a significant milestone in the GPS application, it appears as engineering stakeout, topographic mapping, various control measure has brought a new dawn, a very land outside the industry to improve the operating efficiency.

In this article, we will single baseline RTK combined system performance Compass and GPS systems were compared and analyzed. This comparison is estimated global navigation satellite system parameters and their reliability in the success rate of integer ambiguity measures and accuracy. Reliability is a measure of the robustness of the underlying model, and can be divided into internal and external reliability. Observing the relationship between the internal reliability of a model system to test the ability of the error, and on the outside is the reliability of the estimated parameters in these models, such models are better able to eliminate the error. For integer ambiguity is determined by the LAMBDA method is to give the majority of users to analyze the data GNSS Compass can bring what kind of benefits, whether used alone or when used in combination with GPS.

We begin with an overview of Compass and GPS systems, and describes the model and method for positioning and for integer ambiguity. Then, we describe the reliability measures, draw relevant conclusions. Finally, we summarize and discuss.

2 Data Collection

2.1 Satellite Tracking

The receiver in experiment is UB240-CORS GPS/BeiDou Dual-frequency receiver, the receiver is capable of receiving GPS and BeiDou satellite signals simultaneously, the output frequency is 1 Hz. Antenna mounted in the roof of Shanghai Ocean University, College of Information. Five geostationary satellites are positioned at 58.75° E(C5), 80° E (C4), 110.5°E(C2 yet to launch), 140°E (C1) and 160°E(C4), and 30 non-geostationary orbit (Non-GEO) is consist of 27 medium earth orbit (ME0) satellite and three inclined geosynchronous orbit (IGSO) satellites. The ME0 satellite orbital altitude 21 500 km, orbital inclination of 55°, and evenly distributed in three orbital planes. IGSO satellite orbital altitude 36 000 km, and evenly distributed in three Synchronous orbit inclined plane. Both of GPS satellites and Beidou-2 satellites are using code division multiple access methods, experimental satellite tracking GPS L1 frequency is 1575.42 MHz, L2 frequency is 1227.6 MHz, the BeiDou-2 satellite B1 frequency is 1561.098 MHz, B2 frequency is 1207.14MHz.

2.2 GPS/BeiDou-2 Portfolio Positioning Model

BeiDou satellite navigation system time system at 0:00 on January 1, 2006 as the start, keeping 14 seconds between the system time of GPS (GPS= BeiDou+14s). The pseudorange observation equation of GPs and Beidou-2 is as follows:

$$\min_{b,a} \|y - Bb - Aa\|_{Q_{yy}}^2, b \in R^p, a \in Z^q$$

Which $\|\cdot\|_{Q_{yy}}^2 = (\cdot)^T Q_{yy}^{-1} (\cdot)$, Q_{yy} is the GNSS observer variance-covariance matrix (VCV), R^p is p-dimensional real space, Z^q is q-dimensional integer space. When calculating the position of the satellite, the calculation method of BeiDou-2 MEO and IGS0 satellites is the same with GPS satellite. GE0 taken a special matrix, and it fixed the influence of relativistic effects at the same time. Ionospheric and tropospheric delays are use commonly models Klobuchar and saastamoinen to estimate.

3 Data Analysis

3.1 Number of Visible Satellites

The number of visible satellites is an important indicator of the performance of the positioning. The number of equations involved in solving the location was greater than the number of unknowns only when the receiver receives signals from at least four satellites, then the positioning is possible. So the effectively number of visible satellites is a sign of the effectiveness of positioning.

3.2 Position Dilution of Precision

Positioning matrix can be represented by linearized matrix as:

$$G \begin{pmatrix} \Delta x + \varepsilon_x \\ \Delta y + \varepsilon_y \\ \Delta z + \varepsilon_z \\ \Delta \delta_i + \varepsilon_{\delta i} \end{pmatrix} = b + \varepsilon_p$$

G is the satellite geometry matrix, ε_x , ε_y , ε_z , ε_{δ} represent positioning error in each direction and time respectively. ε_p is the measurement error vector, Δx , Δy , Δz , $\Delta \delta$ represent users displacement vector and the receiver clock error. b is the deviation between the estimated and actual values, (1) can be solved by the least square method. Thereby positioning error can be drawn with the relationship between the measurement error:

$$COV \begin{pmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \\ \varepsilon_{\delta i} \end{pmatrix} = H \sigma_{URE}^2$$

In the formula (2) $H=(GTG)^{-1}$, H is called the coefficient matrix, it is a symmetric matrix of a series of 4×4 . Formula indicates the variance of the measurement error is enlarged into a weighting coefficient matrix positioning error variance. Therefore, the smaller the matrix elements of the measurement value, the lower the degree of error is amplified. Spatial location accuracy factor represents the coefficient matrix in the right position on the amplification factor space.

$$\check{a} = \arg \min_{a \in Z^q} \|\hat{a} - a\|_{Qaa}^2$$

3.3 Accuracy and Precision

The accuracy and precision represents systematic measurement error and random error, the accuracy reflects the measurement results with respect to the true value, and the precision reflects the degree of stability of positioning result. The accuracy can be calculated by subtracting solver results and true position, precision can be calculated by the average RMS of each solver results.

GPS system’s deviation in the horizontal direction is slightly higher than the Integration Beidou-2/GPS, but is better than the Beidou-2. Beidou-2’s positioning error in the horizontal direction and the vertical direction is less than 15 m and 10m, and the deviation in the horizontal direction is undulating. After the comparison of the number of visible satellites and space position accuracy factor, we can found that the peak appear at small number of visible satellites and great spatial accuracy factor period.

Table 1. RMS of positioning in three days

Date	System	Root Mean Square		
		East	North	upper
2013-11-02	combination	0.988	1.293	1.246
	GPS	1.102	1.359	1.579
	Beidou	1.827	2.060	2.256
2013-11-03	combination	0.926	1.142	1.336
	GPS	1.141	1.349	1.618
	Beidou	1.892	1.943	2.221
2013-11-04	combination	1.097	1.119	1.207
	GPS	1.294	1.385	1.618
	Beidou	1.710	2.015	2.254

Table 1 lists RMS of positioning in four days, the smaller the RMS, the more stable results illustrate positioning. After the comparison between the data in the table it can be seen the arrangement of each system RMS in ascending order of Integration Beidou-2/GPS, GPS system and Beidou-2.

Table 1 shows the Integration Beidou-2/GPS has the highest precision, the accuracy of Beidou-2 is lower than GPS. The main reason is that we do the experiment in July, Beidou-2 regional distribution system is not yet complete at that

time, particularly there only two satellites in orbit. Currently, Beidou-2 has launched two satellites in orbit, and the number of satellites in orbit reached 4. However, due to the current two satellites in orbit is still an experimental stage the accuracy assessment can not be achieved, but Beidou-2 positioning accuracy will increase with the improvement of Beidou-2 satellite positioning environment.

4 Conclusion

Through the above analysis, it can be concluded that Integration Beidou-2/GPS positioning mode has more visible satellites than a single mode. the spatial accuracy factor is smaller, the results of positioning are more accurate and stable, and GPS's positioning performance is better than the current Beidou-2. Beidou-2 positioning system in the horizontal and vertical directions can now achieve positioning accuracy of 5 m or less, and 10 m or less.

References

1. Grelier, T., Ghion, A., Dantepal, J., Ries, L., DeLatour, A., Issler, J.-L., Avila-Rodriguez, J., Wallner, S., Hein, G.W.: Compass signal structure and first measurements. In: Proceedings of ION GNSS 2007, Fort Worth, TX, pp. 3015–3024 (2007)
2. Chen, H., Huang, Y., Chiang, K., Yang, M., Rau, R.: The performance comparison between GPs and BeiDou-2/compass: a perspective from Asia. *J. Chin. Inst. Eng.* 32(5), 679–689 (2009)
3. Yang, Y., Li, J., Xu, J., Tang, J., Guo, H., He, H.: Contribution of the Compass satellite navigation system to global PNT users. *Chin. Sci. Bull.* 56(26), 2813–2819 (2011)
4. Zhang, S., Guo, J., Li, B., Rizos, C.: An analysis of satellite visibility and relative positioning precision of COMPASS. In: Proceedings of Symposium for Chinese Professionals in GPS, Shanghai, People's Republic of China, pp. 41–46 (2011)
5. Shi, C., Zhao, Q., Li, M., Tang, W., Hu, Z., Lou, Y., Zhang, H., Niu, X., Liu, J.: Precise orbit determination of Beidou satellites with precise positioning. *Sci. China Earth Sci.* 55, 1079–1086 (2012)
6. Steigenberger, P., Hauschild, A., Montenbruck, O., Rodriguez-Solano, C., Hugentobler, U.: Orbit and clock determination of QZS-1 based on the CONGO network. In: ION-ITM-2012, Newport Beach, California (2012)
7. Odijk, D., Teunissen, P.J.G.: Characterization of between-receiver GPS-Galileo inter-system biases and their effect on mixed ambiguity resolution. *GPS Solutions*, 1–13 (2013), doi:10.1007/s10291-012-0298-0

Texture Image Classification Using Gabor and LBP Feature

Youfu Du

Yangtze University, Hubei, China
dyf@yangtzeu.edu.cn

Abstract. This paper presents a feature fusion based texture image classification method simultaneously using Gabor and Local Binary Patterns (LBP) feature. LBP and Gabor wavelets are two widely used two successful local image representation methods. This paper proposes two kinds of feature fusion methods, which perform in feature level and matching score level, respectively. We show that combining the two successful local image representations, i.e. Gabor wavelets and LBP, gives considerably better performance than either alone. Experiment results on MIT texture database demonstrate the effectiveness of our method.

Keywords: texture image classification, gabor wavelets, LBP.

1 Introduction

The textures exist in natural scenes captured in the image. The image texture is defined as a function of the spatial variation in gray values. The texture analysis is useful in a variety of applications, and it has been a subject of intense study by many researchers [1]. One immediate application of image texture is the recognition of image regions using texture properties. Image textures are one way that can be used to help in segmentation or classification of images. The goal of texture classification then is to produce a classification map of the input image where each uniform textured region is identified with the texture class it belongs to. Texture classification for images has been a hot research topic in computer vision for many years [2]. The texture classification has many potential applications. However, despite many potential areas of application for texture analysis in industry, there is only a limited number of successful examples. A major problem is that textures in the real world are often not uniform, because it changes in orientation, scale or other visual appearance.

Many appearance based approaches have been proposed to deal with texture classification problems. PCA [3] and LDA [4] are two widely used appearance based approaches, which have been the state of the art texture classification techniques. The PCA method extracts the features from the image matrix by projecting the image matrix along the projection axes that are the eigen-vectors of the covariance matrix. As the results, a texture subspace is constructed to represent the texture image. Similarly, LDA constructs a discriminant subspace, which is constructed to

distinguish optimally textures of different subjects. In 2003, an improved PCA technique named two-dimensional PCA (2DPCA) was proposed by Yang et al. [5]. The 2DPCA directly extracts the features from the image matrix by projecting the image matrix along the projection axes that are the eigen-vectors of the 2D images covariance matrix. As the covariance matrix of 2DPCA has a lower dimensionality than that of PCA, 2DPCA is computationally more efficient than PCA. Motivated by 2DPCA, Xu et al. proposed to combine two solution schemes of 2DPCA to extract features from matrixes. Gao and Zhang et al. propose the two-dimensional independent component analysis (2DICA) that directly evaluates the two correlated demixing matrices from the image matrix without matrix-to-vector transformation.

The Gabor wavelets [6,7] based image preprocessing method achieves great success in texture classification. The Gabor wavelets, whose kernels are similar to the response of the two-dimensional receptive field profiles of the mammalian simple cortical cell, exhibit the desirable characteristics of spatial locality and orientation selectivity. The Gabor transformed image is represented by the convolution results of the image matrix with the Gabor wavelet. Since the Gabor features are extracted in local regions, they are less sensitive to variations of illumination than the holistic features.

LBP is a feature extraction method which considers both shape and texture information to represent the images [8]. A straight forward extraction of the feature vector (histogram) is adopted in LBP. The Local Binary Pattern (LBP) features are extracted and concatenated into a single feature histogram efficiently representing the image. The textures of the regions are locally encoded by the LBP patterns. The idea behind using the LBP features is that the texture images can be seen as composition of micro-patterns which are invariant with respect to monotonic grey scale transformations [9]. There is an extension to the original operator, in which it defined the so-called uniform patterns: an LBP is 'uniform' if it contains at most one 0-1 and one 1-0 transition when viewed as a circular bit.

In this paper, we seek to integrate the Gabor feature and LBP feature for texture classification. In texture image classification applications, when we get multiple feature sets of the pattern samples, it is very important to achieve a desirable recognition performance based on the feature sets. Feature fusion technology has been developed rapidly in the past years. There are mainly the following types of feature fusion strategies, which are fusion in data level fusion, fusion in feature level fusion, fusion in matching score level fusion and fusion in decision level fusion. Data fusion simply combines different domains of raw data to form a new raw data, but it is difficult to implement in practice because of the following reasons: the feature sets of multiple modalities may be incompatible. Decision fusion combines multiple classifiers, but it is difficulty to achieve good performance. So we employ the fusion frameworks in feature level and matching score level, which has been applied to texture image classification.

The paper is organized as follows. In the next section we give a review of related work. Section 3 describes our method. In section 4, we give a number of experiment results. Section 5 offers our conclusions.

2 Related Works

2.1 Gabor Transform Method

Gabor filters have been used extensively in image processing, texture analysis for the excellent property of simulating the receptive fields of simple cells in the visual cortex. The Gabor filter is generated from a wavelet expansion of the Gabor kernels, exhibit desirable characteristics of spatial locality and orientation selectivity. Figure 1 gives an example of Gabor filter.

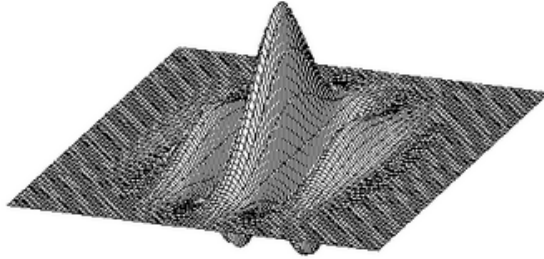


Fig. 1. An example of Gabor filter

The Gabor filter takes the form of a complex plane wave modulated by a Gaussian envelope function. The Gabor filter we used can be formulated in spatial-frequency domain as:

$$\psi_{u,v}(z) = \frac{\|k_{u,v}\|}{\sigma^2} e^{(-\|k_{u,v}\|^2 \|z\|^2 / 2\sigma^2)} [e^{izk_{u,v}} - e^{-\sigma^2 / 2}]$$

where, $z = (x, y)$, $\sigma = 2\pi$, $k_{u,v} = \begin{pmatrix} k_v \cos \phi_u \\ k_v \sin \phi_u \end{pmatrix}$, k_v and ϕ_u controls the scale

and orientation of the Gabor wavelet, respectively. The first term in the brackets of the above Eq. is the oscillatory part of the kernel and the second compensates is the DC value. Let image matrix $I(z)(z = (x, y))$ be a image matrix, and then the Gabor transformed image is represented as the convolution of $I(z)$ with the Gabor wavelet $\psi_{u,v}(z)$, which can be defined as the following equation:

$$O_{u,v}(z) = I(z) * \psi_{u,v}(z)$$

The image matrix $I(z)$ is corresponding to 40 Gabor transformed images ($O_{k_0, \phi_0}(z), O_{k_0, \phi_1}(z), \dots, O_{k_4, \phi_7}$).

2.2 LBP

• LBP Coding

Ojala et al. [9] proposed the LBP operator in 1996. Local Binary Pattern (LBP) features have performed very well in various applications, including texture classification and segmentation. The LBP operator takes a local neighborhood around each pixel, thresholds the pixels of the neighborhood at the value of the central pixel and uses the resulting binary-valued image patch as a local image descriptor. The original LBP operator labels the pixels of an image by thresholding the 3-by-3 neighborhood of each pixel with the center pixel value and considering the result as a binary number. It was originally defined for 3×3 neighborhoods, giving 8 bit codes based on the 8 pixels. The resulting LBP can be expressed in the decimal form as

$$LBP(x_c, y_c) = \sum_{n=0}^7 s(i_n - i_c) 2^n$$

where n runs over the 8 neighbors of the central pixel, i_c and i_n are the gray-level values of the central pixel and the surrounding pixel, and $s(x)$ is 1 if $x > 0$; otherwise, $s(x)$ is 0. Researchers have made an extension of the original operator. The operator was extended to use neighborhood of different sizes, to capture dominant features at different scales. Using circular neighborhoods and interpolating the pixel values allow any radius and number of pixels in the neighborhood. Figure 2 gives an example of LBP coding on a pixel.

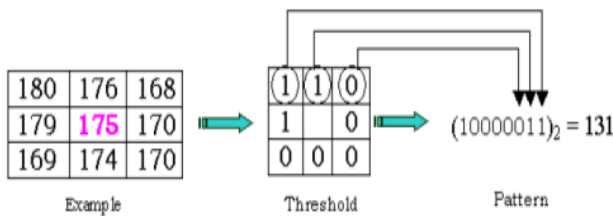


Fig. 2. LBP coding on a pixel

• The Uniform Patterns in LBP

It is noticed that most of the texture information was contained in a small subset of LBP patterns. These patterns, called uniform patterns, contain at most two bitwise 0 to 1 or 1 to 0 transitions (circular binary code). 11111111, 00000110 or 10000111 are for instance uniform patterns. They mainly represent primitive micro-features such as lines, edges, corners. The uniform patterns represent local primitives such as edges and corners. It was observed that most of the texture information was contained in the uniform patterns.

People usually label the patterns which have more than 2 transitions with a single label yields an LBP operator, which produces much less patterns without losing too much information. The following figure shows the 56 uniform patterns, and the two remaining uniform patterns are 11111111 and 00000000.



Fig. 3. The 56 uniform patterns. The black point means ‘1’ and the white point means ‘0’.

3 Our Method

In this section, we propose two feature fusion methods in feature level and matching score level, respectively. The first stage of the two fusion methods is same. They employ the Gabor transform and LBP coding methods to get the features.

For extracting discriminative information as much as possible, a bank of Gabor filters with several orientations and scales is chosen to extract the features from the image. 5 scales and 8 orientations are used in this paper, i.e.: $k_0 = \pi / 2^{0/2}, k_1 = \pi / 2^{1/2}, \dots, k_4 = \pi / 2^{4/2}$ and $\phi_0 = 0\pi/8, \phi_1 = \pi/8, \dots, \phi_7 = 7\pi/8$. We show the real part, imaginary part and the magnitude of the 40 Gabor filters (5 scales and 8 orientations) in Fig. 4:

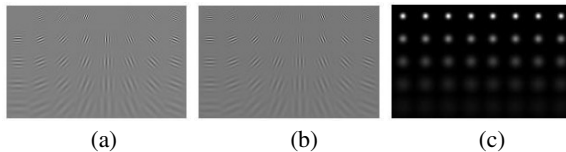


Fig. 4. (a) The real part of the 40 Gabor filters, (b) The imaginary part of the 40 Gabor filters, (c) The magnitude of the 40 Gabor filters

One of the potential problems existing in Gabor feature is that it is redundant and too high-dimensional. For the example: if the size of the image size is 100×80 , the number of the Gabor features will reach $(100 \times 5) \times (80 \times 8)$, which is incredibly large for the following feature extraction and classification method. Additionally, no evidence is found that every Gabor feature dose favor to improve classification accuracy. It is meaningful to conduct feature dimension reduction on all the Gabor features. As each Gabor transformed matrix is corresponding to a Gabor filter, we equate the feature dimension reduction to the selection of Gabor filters. Dozens of Gabor filters are usually adopted for constructing the ensemble Gabor transformed matrix, so exhaustive search is too time-consuming to get the solution. We design a heuristic search algorithm with forward selection. In our algorithm, the sum of the absolute values of the eigen-values of $G_w^{-1}G_b$ is used to evaluate the quality of the selected subset, where G_w, G_b are the between-class and within-class scatter matrices of 2DLDA respectively. Employing this criterion ensures that the new ensemble Gabor transformed matrix has the maximum Fisher's ratio, which is favorable for improving the classification accuracy of the training samples. The Gabor filter selection algorithm follows these steps:

Step 1. Initialize the parameter ν that is the number of the filters to be selected.

Step 2. Select the first Gabor filter ψ_{s_1} from the Gabor filter set $\{\psi_1, \psi_2, \dots, \psi_k\}$ according to the criterion.

Step 3. There are $k - 1$ choices for selecting the second Gabor filter. Denote the Gabor transformed matrix corresponding to ψ_i as O_i . For each choice such as ψ_i , we ensemble the two Gabor transformed images of each training sample as the matrix $[O_{s_1}, O_i]$. Then, we calculate the value of the criterion function for the choice of ψ_i .

ψ_{s_2} is selected as the second optimal Gabor filter, which has the maximum criterion function value among the $k - 1$ choices.

Step 4. When the number of the Gabor filters selected reaches ν , the algorithm terminates, otherwise, go to Step 5.

Step 5. Supposing t Gabor filters has been selected, then there are $k - t$ choices for selecting the $(t + 1)$ th Gabor filter. For each choice such as ψ_i , we ensemble the $t + 1$ Gabor transformed images of each training sample as the matrix $[O_{s_1}, O_{s_2}, \dots, O_{s_t}, O_i]$. Then, we calculate the criterion function for the choice of ψ_i . $\psi_{s_{(t+1)}}$ is selected as the $(t + 1)$ th Gabor filter, which has the maximum criterion function value among the $k - t$ choices.

By the above algorithm, ν optimal Gabor filters $(\psi_{s_1}, \psi_{s_2}, \dots, \psi_{s_\nu})$ are selected. Our optimal ensemble Gabor transformed image is constructed by the form of $[O_{s_1}, O_{s_2}, \dots, O_{s_\nu}]$.

The individual sample image is divided into several small non-overlapping blocks with same size. Histograms of LBP codes are calculated over each block and then concatenated into a single histogram representing the image.

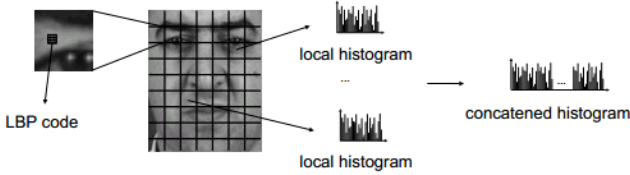


Fig. 5. Block based LBP

For the fusion method in feature level, the key stage is how to effectively combine the features. The simplest combination method is directly merged the Gabor feature vector and the LBP feature vector into one feature vector. The distance e denotes the matching score in our method. If $p = \arg \min_i e(i)$, then the testing sample is classified into the p -th subject. The following figure shows the framework of the fusion method in feature level.

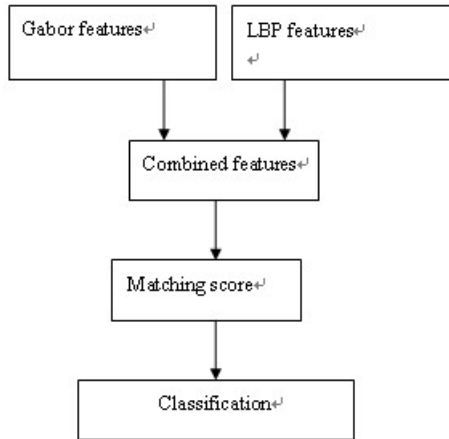


Fig. 6. The framework of the fusion method in feature level

For the fusion method in matching score level, the matching score in the two channels (Gabor feature and LBP) are computed, respectively. After getting the two matching scores e_1, e_2 of the test sample to the training image, Let $e(i) = ce_1(i) + (1 - c)e_2(i)$. If $p = \arg \min_i e(i)$, then the testing sample is classified into the p -th subject.

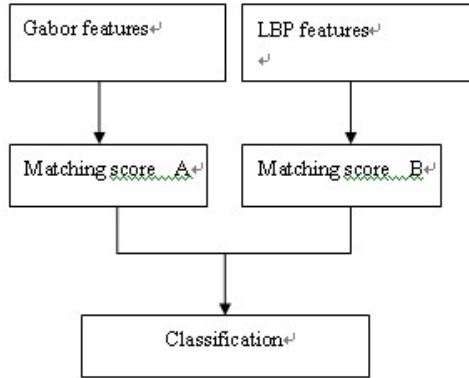


Fig. 7. The framework of the fusion method in Matching score level

4 Experiments

We assess the performance of our method on the MIT texture database, which includes 40 texture images. Figure 8 shows some images from this database. Each image has the same resolution of 512×512 . Figure 8 shows some images from this database. We divided each image into 4 sub images with the size of 256×256 . Figure 9 presents the 4 sub images obtained from the original image.

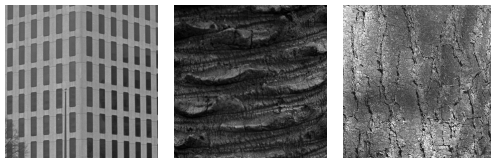


Fig. 8. Some images from this database

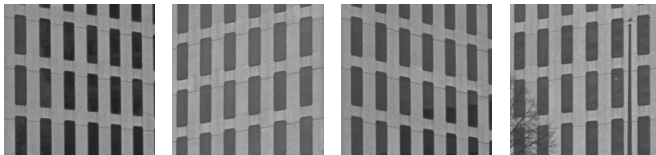


Fig. 9. 4 Sub images obtained from the original image

First, we use the 40 Gabor filters mentioned in section 3 to generate the Gabor feature of each image. Some examples of preprocessed images are shown in figure 10.

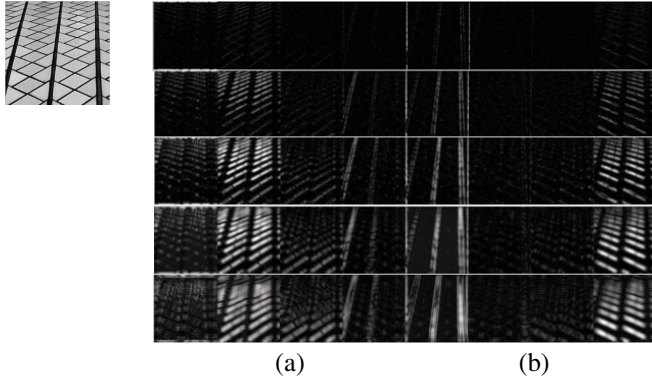


Fig. 10. (a) a sub-image from the MIT texture database (b) the Gabor features of the left image

LBP also has been used for image representation. LBP histograms are extracted from the LBP coding image. It was observed that most of the texture information was contained in the uniform patterns. It has been noted that viewing the non-uniform patterns as one pattern produces much less patterns and does not lose too much information. so uniform patterns was used to reduce the length of LBP histograms. Figure 11 presents some 59-bin histograms obtained from the corresponding LBP coding images.

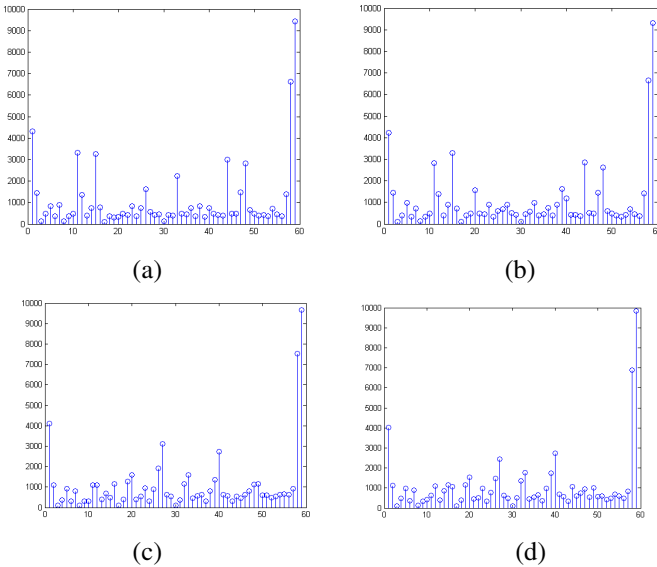


Fig. 11. (a) and (b) are the histograms of the two sub-images with the same texture. (c) and (d) are the histograms of the two sub-images with the same texture

As the comparisons, the state of the art image reorientation methods including original feature, DCT, Gabor wavelets. In the fusion method in the matching score level, we set e_1 be 0.1, 0.2, ..., 0.9. Among all the values of the parameter, we chose the one that has the maximum classification accuracy. The classification accuracies of these methods are presented in the following table.

Table 1. The classification accuracies on MIT texture database

methods	classification accuracy (%)
original feature	20.62
DCT (25 features)	34.38
DCT (100 features)	30.63
DCT (400 features)	26.25
DCT (65536 features)	20.62
Gabor (2621 features)	70.36
Gabor (655 features)	66.25
Fusing LBP and Gabor in feature level	98.75
Fusing LBP and Gabor in matching score level	98.75

5 Conclusions

This paper seeks to integrate the Gabor wavelets and LBP image presentation methods for texture classification. For achieving this aim, we proposed an optimal Gabor representation method and employed two feature fusion methods. The optimal Gabor representation method extracts the most representative and discriminative information from Gabor features. The two feature fusion methods are simple and effective. The results of the experiments carried on MIT texture database show that our method our method has strong ability in classifying texture images.

References

1. http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OJALAL/texclas.htm
2. Varma, M., Zisserman, A.: A Statistical Approach to Texture Classification from Single Images. *International Journal of Computer Vision* 62 (1-2), 61–81
3. Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cognitive Neurosci.* 3(1), 71–86 (1991)
4. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Machine Intell.* 19(7)

5. Yang, J., Zhang, D., Frangi, A.F., Yang, J.Y.: Two dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Trans. Pattern Anal. Machine Intell.* 26(1), 131–137 (2004)
6. Yang, M., Zhang, L.: Gabor Feature based Sparse Representation for Face Recognition with Gabor Occlusion Dictionary. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *ECCV 2010, Part VI. LNCS*, vol. 6316, pp. 448–461. Springer, Heidelberg (2010)
7. Lee, T.S.: Image representation using 2D Gabor Wavelets. *IEEE Transactions on PAMI* 18(10) (1996)
8. Ojala, T., Pietikainen, M., et al.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(7), 971–987 (2002)
9. Ojala, T., Pietikainen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distribution. *Pattern Recognition* 29(1), 51–59 (1996)

Part VI
Multimedia Innovative Computing

Effective Moving Object Detection from Videos Captured by a Moving Camera

Wu-Chih Hu¹, Chao-Ho Chen², Chih-Min Chen², and Tsong-Yi Chen²

¹Department of Computer Science and Information Engineering,
National Penghu University of Science and Technology, Penghu, Taiwan, R.O.C.
wchu@npu.edu.tw

²Department of Electronic Engineering,
National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan, R.O.C.
{thouho, 1098305145, chentso}@cc.kuas.edu.tw

Abstract. This paper presents an effective method to detect moving objects for videos captured by a moving camera. Moving object detection is relatively difficult to videos captured by a moving camera, since in the case of the video filmed by moving cameras, not only do the objects move, but also the frames shift. In the proposed schemes, the feature points in the frames are first found and then classified into the foreground and background. Next, the foreground regions and image difference are obtained and then further merged to obtain moving object contours. Finally, the moving object is detected based on the motion history of the continuous motion contours and refinement schemes. Experimental results show that the proposed method performs well in terms of moving object detection.

Keywords: moving object, moving camera, motion history.

1 Introduction

Because of the increasing demand for safety and security, intelligent surveillance has received a lot of attention. Intelligent surveillance has a wide range of applications [1-6], such as event detection, behavior description, tracking and identification of moving objects, object counting, and security. Moving object detection plays an important role in an intelligent surveillance. Usually, video object segmentation is necessary obtained to detect the moving object.

Many methods have been proposed for video object segmentation. Generally, these methods can be roughly classified into two types [7-11]: background construction-based video object segmentation, and foreground extraction-based video object segmentation. In background construction-based video object segmentation, the background information is first constructed. Then, the video object is obtained based on the difference between the background and the current frame. In foreground extraction-based video object segmentation, temporal information, spatial information, or temporal-spatial information is first used to obtain an initial video object. Then, the video object in the

successive frame can be obtained using motion information, change information, and other feature information.

Background construction-based video object segmentation and foreground extraction-based video object segmentation are suitable used for static background and dynamic background videos captured by a fixed camera, respectively. However, the background, foreground, and camera are all moving objects in the videos captured by a moving camera. In comparison to fixed cameras, it is relatively difficult to videos captured by a moving camera, since in the case of the video filmed by moving cameras, not only do the objects move, but also the frames shift. With the assistance of object segmentation skills, the shapes of the moving objects fail to be effectively segmented and detected. Therefore, the known background construction-based video object segmentation and foreground extraction-based video object segmentation schemes are not suitable used for videos captured by a moving camera to detect moving objects. Video captured by a moving camera has a wide range of applications, such as mobile robot, robot-car, and vehicle video recorder. Therefore, it is a challenging task to obtain accurate moving object detection for videos captured by a moving camera.

Jodoin et al. [12] proposed a robust moving object detection for both fixed and moving camera captured video sequences by using Markov random field (MRF) to obtain label fields fusion. Wang [13] used the capability of MRF model to propose the detecting moving vehicles in different weather conditions, but this approach is limited due to its only applicability in gray scale videos. In order to handle the spatial ambiguities of gray values, Ghosh et al. [14] proposed a novel region matching-based motion estimation scheme to detect objects with accurate boundaries from videos captured by moving camera. However, the computational cost of this approach is high, and it does not yield good results under the moving object with cast shadows or with occlusion/disocclusion.

The present study proposes an effective method to detect moving objects for videos captured by a moving camera. In the proposed schemes, the feature points in the frames are first found and then they are classified into the foreground and background with the assistance of multiple view geometry. Next, the foreground regions are obtained based on the foreground feature points. The image difference is calculated using the assistance of affine transformation based on the background feature points. Then, moving object contours are obtained by merging the foreground regions and image difference. Finally, the motion history of consecutive frames and refinement schemes are used to obtain the moving object detection.

The rest of this paper is organized as follows. The proposed moving object detection is described in Section 2. Section 3 presents experimental results and evaluations. Finally, the conclusion is given in Section 4.

2 Proposed Moving Object Detection

The proposed moving object detection has three parts: The first part is the classification of feature points. The second part is detection of moving object contours. The third part is the location of moving objects. A flowchart of the algorithm is shown in Fig. 1.

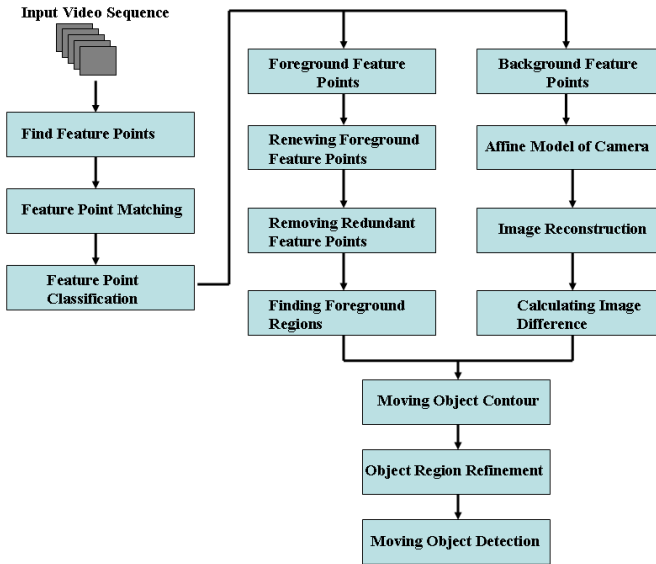


Fig. 1. Flowchart of the proposed method

2.1 Feature Point Classification

In the videos captured by a moving camera, the obtained frame is a global movement. It is very difficult to classify the pixel belonging to the moving objects in the frame with global movement. The magnitude and direction of pixels are usually used to solve the above problem, but it is very easy to have errors because the speed and direction of the moving object are changed at any time.

In this study, the corners are used to raise the accuracy of the movement of pixels. Sobel operation is used to obtain edge detection. Next, Harris corner detector is used to obtain the corners. These obtained corners are used as the feature points, as shown in Fig. 2, where red-cross is the feature point in Fig. 2(b).

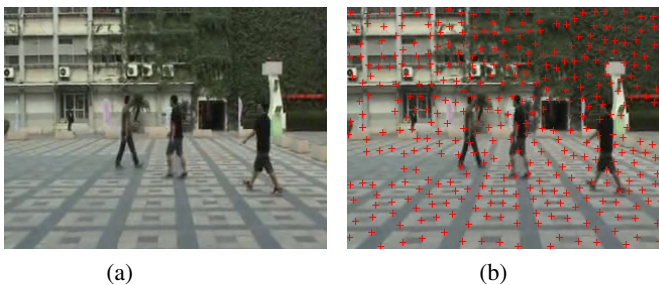


Fig. 2. Corner Detection: (a) Original frame; (b) obtained corners

In the feature point matching, the feature points in two consecutive frames are checked by using the optical flow to evaluate the motion vector and then obtain the appropriate matching result. Next, the Epipolar geometry is used to describe the projective geometry relation of images captured by two cameras with the relative position. Epipolar geometry is used to find the corresponding relation of the points in two images, and it is independent to the size, color, or shape of the object. The fundamental matrix can be used to describe the Epipolar geometry.

Suppose the images captured by two cameras with the relative position as the two consecutive frames. The view geometry can be used to obtain the feature point classification, as shown in Fig. 3, where Fig. 3(a) and Fig. 3(b) are the geometry relation of the pixel in the background and foreground, respectively.

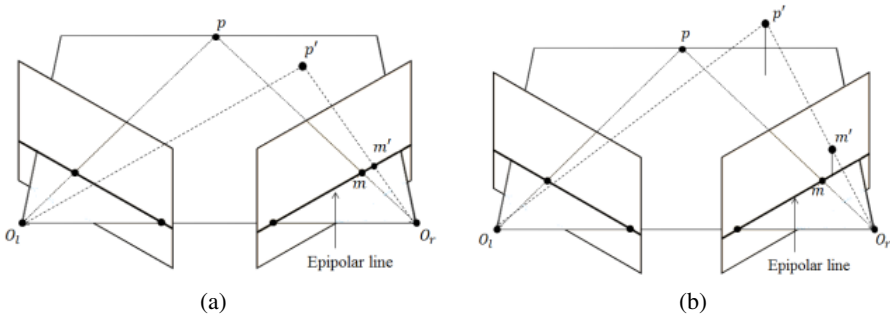


Fig. 3. View Geometry: (a) Background case; (b) foreground case

The fundamental matrix can be calculated as below. Suppose known feature point P_1 and P_2 in two images, respectively. F is the fundamental matrix between two images.

$$P_1 = FP_2 \tag{1}$$

Expand Eq. (1) to obtain Eq. (2).

$$\begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & m_8 \end{bmatrix} \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} \tag{2}$$

Normalize Eq. (2) to one.

$$w \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = \begin{bmatrix} m'_0 & m'_1 & m'_2 \\ m'_3 & m'_4 & m'_5 \\ m'_6 & m'_7 & 1 \end{bmatrix} \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} \tag{3}$$

Expand Eq. (3) to obtain Eq. (4).

$$\begin{cases} m'_0 u_2 + m'_1 v_2 + m'_2 = w u_1 \\ m'_3 u_2 + m'_4 v_2 + m'_5 = w v_1 \\ m'_6 u_2 + m'_7 v_2 + 1 = w \end{cases} \tag{4}$$

Eq. (4) can further be rewritten as Eq. (5).

$$\begin{cases} m'_0u_2 + m'_1v_2 + m'_2 - m'_6u_2u_1 - m'_7v_2u_1 = u_1 \\ m'_3u_2 + m'_4v_2 + m'_5 - m'_6u_2v_1 - m'_7v_2u_1 = v_1 \end{cases} \quad (5)$$

Because 8 unknown variables in Eq. (5) need to be solved, 4 pairs of corresponding feature points are used to solve these variables. To obtain reliable fundamental matrix, n pairs of corresponding feature points and least-squares are used to solve these 8 parameters of fundamental matrix. The Epipolar lines can be drawn using the obtained corresponding feature points and fundamental matrix, as shown in Fig. 4(a), where Fig. 4(a) is the obtained result from Fig. 2(b).

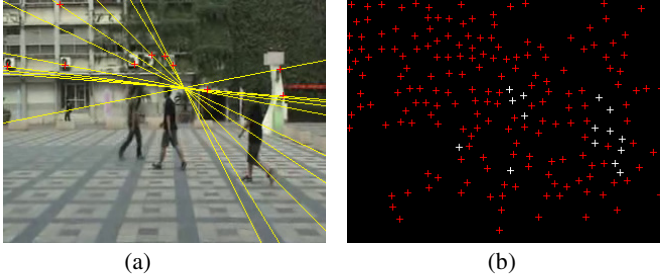


Fig. 4. Feature point classification: (a) Epipolar lines; (b) classification result

Because the fundamental matrix is evaluated using obtained most of feature points which are belonging to the background points, the feature points of the moving objects are not on the Epipolar lines. Therefore, the orthogonal distance between the feature point and the Epipolar line is used to obtain feature point classification. If the orthogonal distance between feature point and the Epipolar line is larger than given threshold, then it is the foreground feature point; otherwise it is the background feature point. Fig. 4(b) is the result using the proposed feature point classification from Fig. 2(b), where red-cross is the background point and white-cross is the foreground point.

2.2 Detection of Moving Object Contours

Once the feature point classification is obtained, a renew scheme of foreground feature points is proposed to raise the reliability of feature point classification. The foreground feature points obtained in the $(k-1)^{th}$ frame are extracted. Next, these extracted foreground feature points are used to match with the feature points in the k^{th} frame to obtain additional foreground feature points. Finally, these additional foreground feature points are merged with the foreground feature points in the k^{th} frame to obtain the renewed foreground feature points, as shown in Fig. 5. Fig. 5(b) is the foreground feature points of Fig. 5(a); Fig. 5(c) is the obtained additional foreground feature points; and Fig. 5(d) is the renewed foreground feature points.

The removing redundant feature point is used for the obtained renewed foreground feature points. The foreground feature point is checked using the block with $k \times k$ pixels centered by this feature point. If there is no another foreground feature point in

this block, then this foreground feature point is removed. Finally, the foreground regions are obtained by merging with all blocks formed by $k \times k$ pixels centered by the foreground feature points.

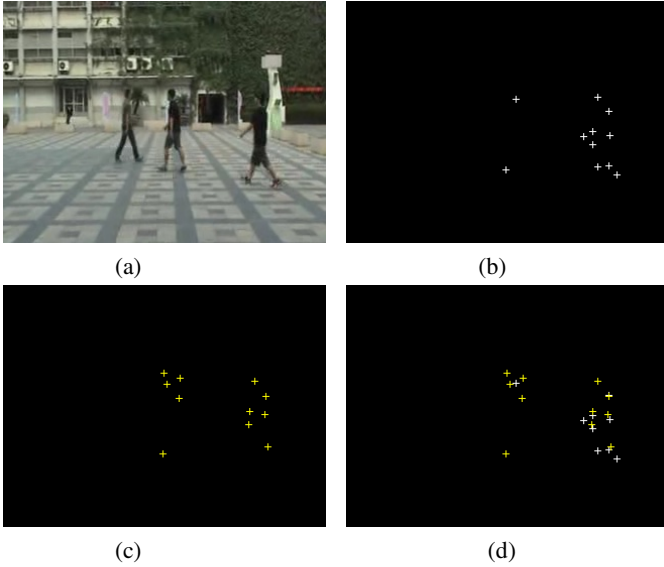


Fig. 5. Result using renewing scheme: (a) Original frame; (b) foreground feature points; (c) additional foreground feature points; (d) renewed foreground feature points

The affine model of the camera is built based on obtained background feature points. The background feature points are used to calculate the optical flow between the corresponding points in the consecutive frames. The obtained information is used to predict the movement of the next frame to further construct the affine model of the camera. Eq. (6) is the affine model of the camera, where (a_0, a_1, a_3, a_4) are the rotation and scaling parameters, respectively; (a_2, a_5) are the translation parameters.

$$\begin{bmatrix} f_x^t \\ f_y^t \end{bmatrix} = \begin{bmatrix} a_0 f_x^{t-1} + a_1 f_y^{t-1} + a_2 \\ a_3 f_x^{t-1} + a_4 f_y^{t-1} + a_5 \end{bmatrix} \tag{6}$$

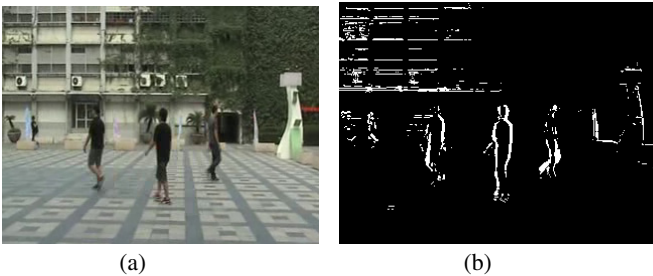


Fig. 6. Result using image difference scheme: (a) Original frame; (b) image difference

The image difference between current frame and one transformed using affine model is calculated by Eq. (7). Fig. 6 is the obtained result of image difference.

$$BI(x, y, t) = \begin{cases} 255 & , \text{if } |frame(x, y, t) - frame(x, y, t')| > T \\ 0 & , \text{otherwise} \end{cases} \quad (7)$$

Finally, the foreground regions and obtained image difference are merged using “AND” operator to obtain the moving object contours, as shown in Fig. 7.

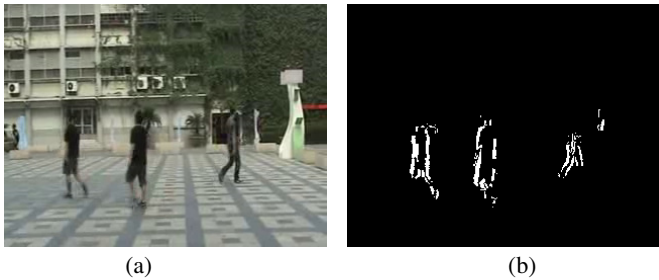


Fig. 7. Detection of moving object contours: (a) Original frame; (b) obtained contours

2.3 Moving Object Location

For compensating the weak regions of moving objects obtained from single frame, the motion history of the continuous motion contours obtained from consecutive three frames is used to increase the regions of moving objects.

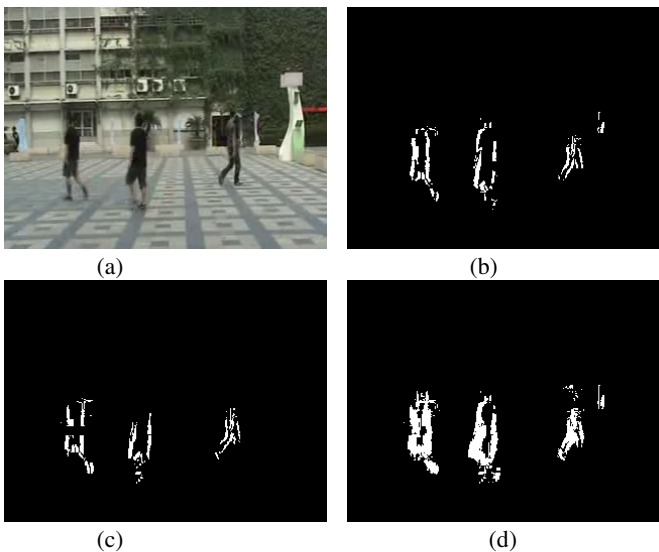


Fig. 8. Region compensating scheme: (a) Original frame; (b) regions of moving objects in the $(k-2)^{th}$ frame; (c) regions of moving objects in the $(k-1)^{th}$ frame; (d) compensated regions of moving objects in the k^{th} frame

Fig. 8 is the obtained result using compensating scheme, where Fig. 8(b) and 8(c) are the regions of moving objects in the $(k-2)^{th}$ and $(k-1)^{th}$ frames, respectively; Fig. 8(d) is the obtained compensated regions of moving objects in the k^{th} frame.

Object region refinement is applied to improve moving object region. The whole image is first processed using down-sample scheme and the morphology processing is further used. The dilation operator is used three times and then the erosion operator is further used three times to obtain more complete moving object regions. Next, the up-sample is used to obtain original size of the image and binarization processing is further applied. Then, fast 4-connected component labeling [5] is applied on the obtained moving object regions to remove the small isolated regions, as shown in Fig. 9(a). Finally, the minimum bounding box is used to mark the moving object to obtain the moving object detection, as shown Fig. 9(b).

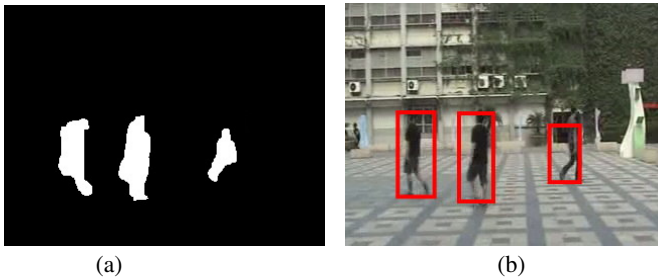


Fig. 9. Moving object detection: (a) Object region refinement; (b) obtained results

3 Experimental Results

Experiments were conducted on a computer with an Intel Core i5 2.8 GHz CPU and 4GB of RAM. The algorithms were implemented in Visual Studio C++. Two tested video were used to evaluate the performance of the proposed method. The sequence of building-penthouse platform consists of 339 frames, each with a size of 320×240 pixels. Fig. 10 is the obtained results using the proposed method. This sequence has total 625 moving objects, and 556 moving objects were detected. Therefore, the accuracy rate is 88.96% for the sequence of building-penthouse platform. The piazza sequence consists of 474 frames, each with a size of 320×240 pixels. Fig. 11 is the obtained results using the proposed method. This sequence has total 726 moving objects, and 602 moving objects were detected. Therefore, the accuracy rate is 82.92% for the piazza sequence. The average accuracy rate is 85.71%. The experimental results showed that the proposed method has good performance in the moving object detection from videos captured by a moving camera.

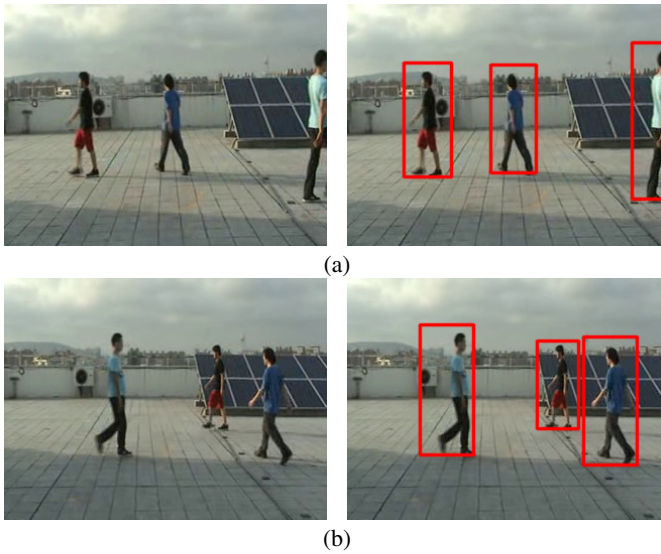


Fig. 10. Obtained results of building-penthouse platform sequence: (a) The 147th frame; (b) the 225th frame

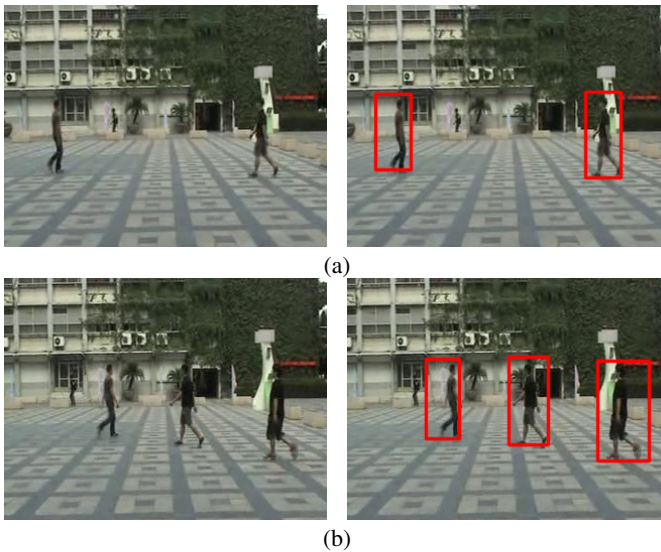


Fig. 11. Obtained results of piazza sequence: (a) The 131th frame; (b) the 178th frame

4 Conclusion

An effective method to detect moving objects for videos captured by a moving camera was proposed. Feature points in the frames are first detected and classified

into the foreground and background. Next, the foreground regions and image difference are merged to obtain moving object contours. Finally, motion history of consecutive frames and object region refinement are used to obtain the moving object detection. Objective evaluation results show that the average accuracy rate is 85.71% for the tested video sequences. Therefore, the proposed method has good performance in moving object detection from videos captured by a moving camera.

Acknowledgments. This paper has been supported by the National Science Council, Taiwan, under grant no. NSC102-2221-E-346-007 and NSC102-2622-E-151-011-CC3.

References

1. Huang, D.-Y., Chen, C.-H., Hu, W.-C., Yi, S.-C., Lin, Y.-F.: Feature-based Vehicle Flow Analysis and Counting for a Real-Time Traffic Surveillance System. *Journal of Information Hiding and Multimedia Signal Processing* 3(3), 282–296 (2012)
2. Dupuis, Y., Savatier, X., Ertaud, J.-Y., Vasseur, P.: Robust Radial Face Detection for Omnidirectional Vision. *IEEE Transactions on Image Processing* 22(5), 1808–1821 (2013)
3. Sugandi, B., Kim, H., Tan, J.K., Ishikawa, S.: Real Time Tracking and Identification of Moving Persons by Using a Camera in Outdoor Environment. *International Journal of Innovative Computing, Information and Control* 5(5), 1179–1188 (2009)
4. Huang, D.-Y., Lin, T.-W., Hu, W.-C., Cheng, C.-H.: Gait Recognition based on Gabor Wavelets and Modified Gait Energy Image for Human Identification. *Journal of Electronic Imaging* 22(4), 043039(1)–043039(11) (2013)
5. Hu, W.-C., Yang, C.-Y., Huang, D.-Y.: Robust Real-time Ship Detection and Tracking for Visual Surveillance of Cage Aquaculture. *Journal of Visual Communication and Image Representation* 22(6), 543–556 (2011)
6. Tian, Y.L., Feris, R.S., Haowei, L., Hampapur, A., Sun, M.-T.: Robust Detection of Abandoned and Removed Objects in Complex Surveillance Videos. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 41(5), 565–576 (2011)
7. Lee, D.-S.: Effective Gaussian Mixture Learning for Video Background Subtraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(5), 827–832 (2005)
8. Wang, L., Yung, N.H.C.: Extraction of Moving Objects from their Background based on Multiple Adaptive Thresholds and Boundary Evaluation. *IEEE Transactions on Intelligent Transportation Systems* 11, 40–51 (2010)
9. Zhu, S., Guo, Z.: An Overview of Video Object Segmentation. In: *Proceedings of International Conference on Industrial Control and Electronics Engineering*, pp. 1019–1021 (2012)
10. Carmona, E.J., Martínez-Cantos, J., Mira, J.: A New Video Segmentation Method of Moving Objects based on Blob-level Knowledge. *Pattern Recognition Letters* 29(3), 272–285 (2008)
11. Hu, W.-C., Chen, C.-H., Huang, D.-Y., Ye, Y.-T.: Video Object Segmentation in Rainy Situations based on Difference Scheme with Object Structure and Color Analysis. *Journal of Visual Communication and Image Representation* 23(2), 303–312 (2012)

12. Jodoin, P.M., Mignotte, M., Rosenberger, C.: Segmentation Framework based on Label Field Fusion. *IEEE Transactions on Image Processing* 16(10), 2535–2550 (2007)
13. Wang, Y.: Joint Random Field Model for All-weather Moving Vehicle Detection. *IEEE Transactions on Image Processing* 19(9), 2491–2501 (2010)
14. Ghosh, A., Subudhi, B.N., Ghosh, S.: Object Detection from Videos Captured by Moving Camera by Fuzzy Edge Incorporated Markov Random Field and Local Histogram Matching. *IEEE Transactions on Circuits and Systems for Video Technology* 22(8), 1127–1135 (2012)

Roadside Unit Deployment Based on Traffic Information in VANETs

Ji-Han Jiang, Shih-Chieh Shie, and Jr-Yung Tsai

Department of Computer Science and Information Engineering,
National Formosa University, Yunlin, Taiwan
{jhjiang,scshie}@nfu.edu.tw, wemee7012@gmail.com

Abstract. In this paper, based on the Vehicle-Assisted Data Delivery (VADD) routing algorithm [3] we present a new approach for the deployment of roadside units (RSU) to improve the data delivery delay in VANET. The main concept of our approach is to add RSUs in intersections that can effectively improve the packet delivery delay. We will address the problem “How to decide which intersection needs to deploy RSU as a data buffer?”

The packet will buffer in a RSU and wait for a vehicle to carry it to the next hop. The RSU increases the opportunity to use wireless communication, decreases the chance to use carry and forwarding.

We have set up a simulation scenario and various traffic conditions to evaluate the performance. The simulation results show that the packet delivery ratio of proposed method has better performance than VADD about 5-10%. The delivery delay of our approach has outperformed than VADD about 15-20% on delivery delay.

Keywords: Vehicular Networks (VANETs), Roadside Unit, Vehicle-Assisted Data Delivery (VADD), Carry-and-Forward, Wireless Sensor Networks.

1 Introduction

In vehicular ad hoc networks (VANETs), packets are delivered by wireless transmission or carry and forwarding [1-8]. The difference is the delay time, transmitted by carry and forwarding is very slowly then wireless transmission. In addition, there are different traffic flows and average vehicle velocities on each road. Some roads are suitable for wireless communication, but others are suitable to carry and forwarding. The key point is the intersection, the packets are delivered to the next road by crossed the intersection.

In this paper, there are two types of the roads (dense and sparse), one is tend to be transmitted by wireless and the other one is tend to be transmitted by carry and forwarding. Based on what types of roads are connected to the intersection. But when an intersection is connected by the two types of roads at the same time, this intersection decides the packet will transmit to which type of roads, fast or slowly. For this reason, we add roadside unit (RSU) at those intersections, based on the

Vehicle-Assisted Data Delivery (VADD) routing algorithm [3], when the packet reaches intersection, but there are no vehicles on the best direction, RSU will store the packet for a while, and waiting for an opportunity. The RSU increases the chance to use wireless communication, decreases the chance to use carry and forwarding. In addition, based on vehicle or RSU, straight or intersection, we provide the algorithms for each conditions, the packet delivery delay is decreased. Finally, we will set up a simulation scenario and various traffic conditions to evaluate the performance.

The rest of this paper is organized as follows. The proposed approach roadside unit deployment based on traffic information is described in Section 2. Section 3 presents experimental results and evaluations. Finally, the conclusion is given in Section 4.

2 Roadside Unit Deployment Based on Traffic Information

There are two kinds of the roads, one is tend to be transmitted by wireless and the other one is tend to be transmitted by carry and forwarding. Based on what kinds of roads are connected to the intersection. But when an intersection is connected by the two kinds of roads at the same time, this intersection decides the packet will transmit to which kind of roads, fast or slowly. For this reason, we add roadside unit (RSU) at those intersections, based on the VADD [3] routing protocol, when the packet reaches intersection, but there are no vehicles on the best direction, RSU will store the packet for a while, and waiting for an opportunity. The RSU increases the chance to use wireless communication, decreases the chance to use carry and forwarding.

2.1 System Concept

(1) Packet Delivery Delay in a Straightway

VADD [3] is based on the idea of carry and forward. The main concept is to select a forwarding path with the smallest packet delivery delay. There are two cases of the packet delivery delay in a straightway. The Case 1 is wireless transmission time and the Case 2 is vehicles carry time (carry and forward).

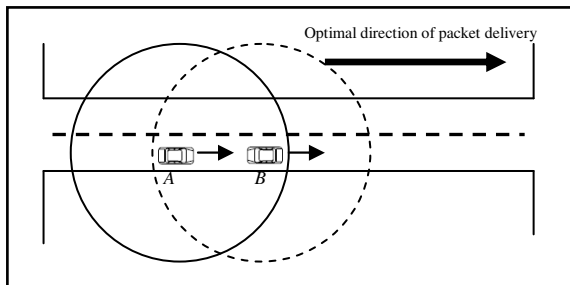


Fig. 1. Select the next vehicle to forward the packet in the straightway

In this scenario, there are two vehicles A and B in a straightway in Figure 1. The solid line and dashed line circles are the wireless transmission range of A and B, respectively. Firstly, we suppose vehicle A is the packet carrier. A will forward the packets to B that is within communication range (called *contacts*) available at the straightway. The packet delivery delay is resulted by wireless transmission (Case 1). Now vehicle B is the packet carrier. There is no *contact* available. Vehicle B will carry the packet continuously and looks for the next forwarding opportunity in the future. The packet delivery delay is resulted by vehicles carry (Case 2).

We can conclude that the main factor of packet delivery delay is carried and forward (Case 2). Our approach is to find a solution that the packet will be forwarded to the one on the road with the smaller delay. Therefore, the main objective of this paper is to reduce the occurrence of Case 2.

(2) Packet Delivery Delay in an Intersection

In above subsection, we find that the main cause of packet delivery delay in a straightway is carry and forward (Case 2). In the VADD algorithm, the packet will be forwarded to the one on the road with the smaller delay. It will try to adopt wireless transmission as soon as possible. The efficiency of VADD is limited by the traffic pattern and the road layout. Based on the existing traffic pattern, a vehicle can find the next hop to forward the packet to reduce the delay.

In VANET, the packet carrier passes the intersection with the packet, and looks for the next forwarding opportunity. The traffic pattern may be dense or sparse. On a dense road, the packets are forwarded hop by hop with wireless transmission. On a sparse road, the packets are carried by a vehicle. Therefore, the packets will be forwarded to a dense or sparse road depending on the selection of next traveling road.

As shown in Figure 2, source *S* has a packet to forward to certain destination *D*. Assume the packet has two choices on selecting the next hop to pass through intersection I_A . These two available paths are P_1 : from I_A moving *south-east* to *D* (dashed line P_1), and P_2 : from I_A moving *east-south-west* to D (solid line), respectively. If there is no other contact available in the front of P_2 to carry the packet to *D*, the VADD algorithm will select shorter path P_1 . However, P_1 is composed of sparse roads. It will greatly increase the packet delivery delay. In this case, we can

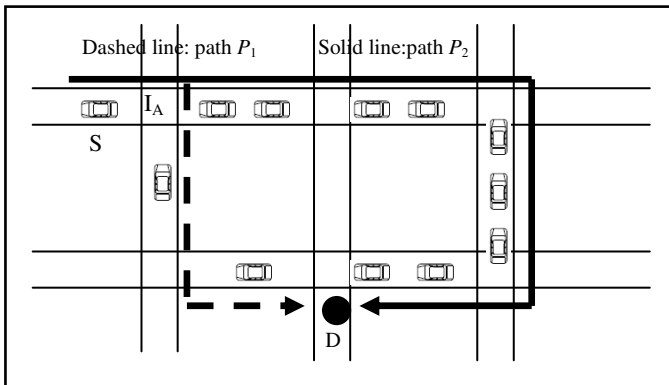


Fig. 2. Select the next vehicle to forward the packet in an intersection

add a roadside unit (RSU) at intersection I_A to play the role of data buffer. The packet will buffer in RSU and wait for a vehicle to carry it to the next hop.

2.2 Roadside Unit Deployment

(1) The Road Types







In VANET, the traffic pattern may be dense or sparse on a road. In this paper, we will define two types of roads according to the density of vehicles on a section of road. The packets are forwarded hop by hop with wireless transmission on a dense road. The packets are carried by a vehicle on a sparse road.

In this paper, we will compute the average traffic flow F'_{ij} of r_{ij} . We define a *critical traffic flow* $C_{ij} = F'_{ij} * \alpha$, where $\alpha=1, 1.5, \text{ or } 2$. Let F_{ij} be the physical traffic flow of r_{ij} . If $(F_{ij} \geq C_{ij})$, r_{ij} is a dense road, the packets are trended to forward hop by hop with wireless transmission in most cases. Otherwise, if $F_{ij} < C_{ij}$, r_{ij} is a spare road, the packets are trended to carry and forward by a vehicle.

(2) The Intersection Types

Suppose that an intersection is connected by four sections of road. There are two types of roads dense and sparse. Therefore, the possible combinations of all traffic patterns are $2^4=16$ kinds. Since the intersection patterns are symmetry, we can rotate and reverse the patterns. As shown in Table 1, there are totally 6 types of intersection patterns. Here, a road with solid line denotes a dense road and a road with solid line denotes a spare road.

Table 1. The Road Types

Types	Intersection Patterns	Discription
T_A		An intersection with heavy traffic load that is connected by four sections of dense road. This is a traffic bottle neck. May be in the town centre.
T_B		An intersection with light traffic load that is connected by four sections of sparse road. May be on the outskirts of town.
T_C		An intersection of a dense road (the main line) and a sparse road.
T_D		May be going into a town centre.
T_E		May be going into a town centre.
T_F		May be going into a traffic jam road.

(3) Location Selection for RSUs

The main concept of our approach is to add RSUs in intersections that can effectively improve the packet delivery delay. We will address the problem “How to decide which intersection needs to deploy RSU as a data buffer?” The packet will buffer in RSU and wait for a vehicle to carry it to the next hop. Let x denote the distance between two vehicles. If ($x \leq R$) the packets are forwarded hop by hop with wireless transmission among vehicles, where R is the wireless transmission range of vehicles. Let σ denote the density of vehicle on a road. The probability of delivery packet hop by hop with wireless transmission is:

$$P(x \leq R) = 1 - e^{-R\sigma} \tag{5}$$

Otherwise, if ($x \geq R$) the packets are carried and forward by a vehicle. The probability of adapting vehicle-assisted data delivery is:

$$P(x > R) = e^{-R\sigma} \tag{6}$$

Firstly, consider the type of intersection T_A of Table 1. This is an intersection with heavy traffic load that is connected by four sections of dense road, where $F_{ij} \geq C_{ij}$. The packets are almost transmitted hop by hop with wireless transmission. That is an intersection with $P(x \leq R) \approx 1$ and $P(x > R) \approx 0$. If we deploy a RSU in intersection of type T_A , to buffer packets, it had to pay the penalties for packet delay. The extra packet delay time is $\frac{1}{F_{ij}}$ in average. So it is not necessary to deploy a RSU in

intersection of type T_A .

Secondly, consider the type of intersection T_B of Table 1. An intersection with light traffic load is connected by four sections of sparse road, where $F_{ij} \leq C_{ij}$. The packets are carried and forward by a vehicle. In this case, using RSU cannot instead of vehicle-assisted data delivery by wireless transmission. Therefore, it is also not necessary to deploy a RSU in intersection of type T_B .

As shown in Table 1, the types: T_C , T_D , T_E , and T_F are intersections of roads with $F_{ij} \geq C_{ij}$ and $F_{ij} < C_{ij}$, respectively. Let's study the type T_C as shown in Figure 3. It is an intersection of mixed kind of roads (i.e. a dense road and a sparse road).

VADD algorithm tends to rout a packet to a road with traffic flow $F_{ij} \geq C_{ij}$. Assume the optimal delivery direction is r_{ab} . However, if the packet is forwarded to r_{ac} , may be unexpected no *contact* exist. Since the traffic flow of r_{ac} is $F_{ij} < C_{ij}$, the packet must be carried and forward by a vehicle. The resulted packet delay is $P(x > R) \times (\frac{L_{ac}}{V_{ac}}) = e^{-R\sigma} (\frac{L_{ac}}{V_{ac}})$. If we deploy a RSU in intersection of type T_C , to buffer

packets, the extra packet delay time is $\frac{1}{F_{ab}}$ in average. So it is better to deploy a RSU

in intersection of type T_C . The intersections of types: T_C , T_D , T_E , and T_F are the same scenarios. Therefore, in this paper we will deploy RSUs in these types of intersections to reduce the packet delay.

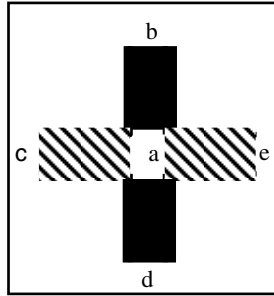


Fig. 3. An intersection of type T_C (an intersection of mixed kind of roads)

If we want to deploy the RSUs in a VANET, initially, collects the t average velocity v_{ij} and traffic load F'_{ij} for each road r_{ij} . Secondly, estimate the optimal value of α for *critical traffic flow* computation ($C_{ij} = F'_{ij} * \alpha$). Determine the type of each intersection accordingly. Finally, deploy the RSUs in intersections of mixed kind of roads according to the trade-off between cost and packet delay.

2.3 Routing Algorithm

In this paper, we use VADD routing algorithm to rout packets in a VANET. We adopt the delay model of VADD compute the optimal direction of packet delivery. For a selected direction, the packet carrier chooses the next intersection towards the selected direction as the target intersection, and applies GPRS algorithm to pass the packet. The packets are *geographical greedy forwarding* towards the target intersection. If the current packet carrier cannot find any *contact* to the target intersection, it buffers the packet in a RSU. If there is no deploying a RSU in this intersection, it chooses the direction with the next lower priority and re-starts the geographical greedy forwarding towards the new target intersection. This process continues until the selected direction has lower priority than the packet carrier's current moving direction. At this time, the packet carrier will continue carrying the packet.

3 Experimental Results

In this section, we evaluate the performance of proposed method (RUDTI). We compare the performance of the proposed approach to two existing approaches: the VADD [3] and GPSR [1].

3.1 The Simulation Environments

The experiment is based on a 600m \times 300m rectangle street area, which presents a grid layout as shown in Figure 4. The street layout is derived and normalized from a snapshot of a real street map in Google Map [9].

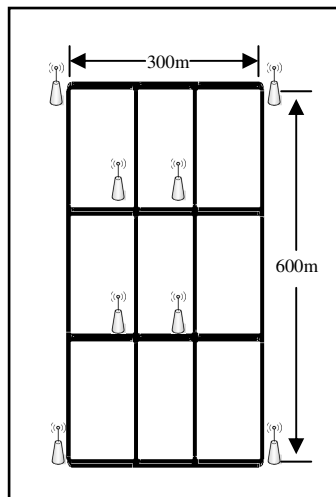


Fig. 4. A grid layout of the simulation street area

In our simulation scenario, the average speed ranges from 30 to 60 Kilo-meters per hour. Different number of vehicles is deployed to the map, and the initial distribution follows the predefined traffic density. To evaluate the performance on different data transmission density, we vary the data sending rate (CBR rate) from 1 to 10 packets per second. The simulation setup is shown in Table 2.

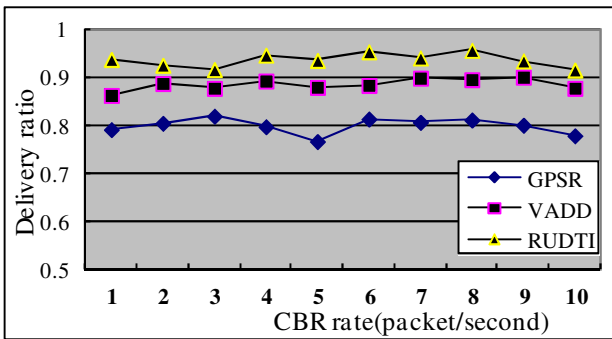
We set up a scenario to simulation and various traffic conditions using QualNet network simulator [10], SUMO [11], and MOVE [6]. The simulation results show that the packet delivery ratio and delivery delay of our approach has a better performance.

Table 2. Simulation setup

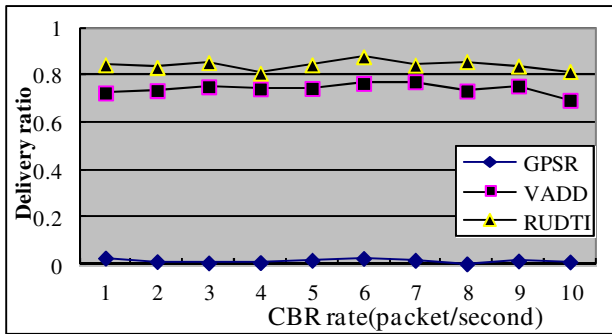
Parameter	Value
Simulation area	600m × 300m
Number of intersections	16
Number of RSUs	8
Number of vehicles	30 -60
Vehicle velocity	0 – 60 Km/hour
Data sending rate (CBR rate)	1 – 10 packets/Sec.
Data packet size	127 Byte
Vehicle beacon interval	0.5 Sec.
Time to life (TTL)	128 Sec.

3.2 Simulation Results

The performance of the approach is measured by the data delivery ratio, the data delivery delay, and the generated traffic overhead. The simulation time is 240 seconds.



(a) 60 vehicles (Dense)



(b) 30 (Sparse)

Fig. 5. Data delivery ratio for CBR

As shown in Figure 5 & 6, the proposed approach RUDTI outperformed VADD for light traffic load (sparse). Our simulation results show that the packet delivery ratio of proposed method has better performance than VADD about 5-10%.

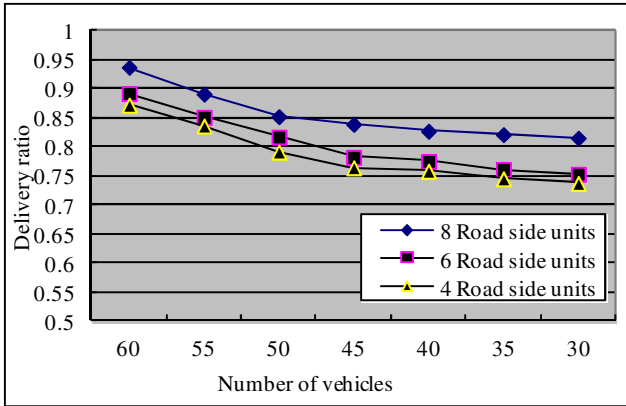


Fig. 6. Data delivery ratio for number of RSUs

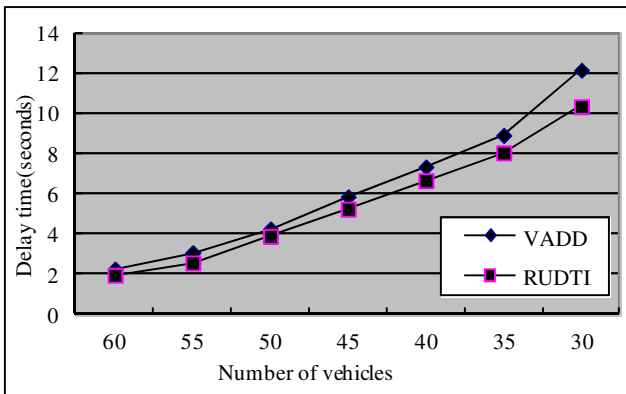


Fig. 7. Data delivery delay for number of vehicles

As shown in Figure 7, the delivery delay will increase for light traffic load. The RUDTI achieves better performance than VADD for the assistant of RSU added routing. Figure 8 show that the RSU increases the chance to use wireless communication, decreases the chance to use carry and forwarding.

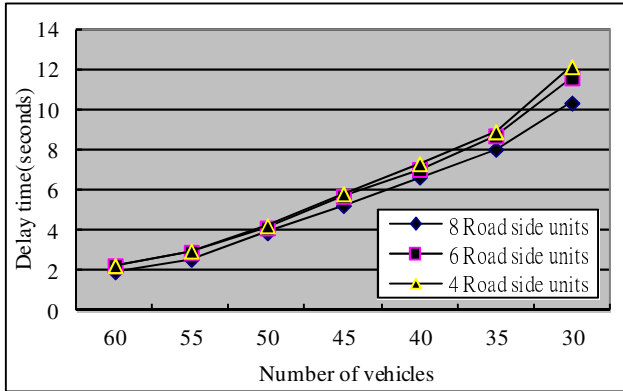


Fig. 8. Data delivery delay for number of RSUs

4 Conclusion

Based on the VADD routing algorithm, we deployed roadside unit (RSU) at intersections to add the packet delivery in VANET. When the packet reaches intersection, but there are no vehicles on the best direction, RSU will store the packet for a while, and waiting for an opportunity. The RSU increases the chance to use wireless communication, decreases the chance to use carry and forwarding.

Our simulation results show that the packet delivery ratio and delivery delay of proposed method has better performance than VADD about 5-10% on delivery ratio and 15-20% on delivery delay, respectively.

Acknowledgments. This work was supported by the Research Grant contract NSC 102-2221-E-150-064 from the National Science Council, Taiwan.

References

1. Karp, B., Kung, H.T.: GPSR: Greedy Perimeter Stateless Routing for Wireless Networks. In: *MobiCom 2000: the 6th Annual International Conference on Mobile Computing and Networking*, pp. 243–254 (2000)
2. Vahdat, A., Becker, D.: Epidemic Routing for Partially-Connected Ad Hoc Networks. Technical Report CS-200006 (2000)
3. Zhao, J., Cao, G.: VADD: Vehicle-Assisted Data Delivery in Vehicular Ad Hoc Networks. In: *IEEE INFOCOM 2006*, pp. 1–12 (2006)
4. Dikaiakos, M.D., Florides, A., Nadeem, T., Iftode, L.: Location-Aware Services over Vehicular Ad-Hoc Networks Using Car-to-Car Communication. *IEEE Journal on Selected Areas in Communications* 25, 1590–1602 (2007)
5. Ding, Y., Wang, C., Xiao, L.: A Static-Node Assisted Adaptive Routing Protocol in Vehicular Networks. In: *ACM International Workshop on Vehicular Ad Hoc Networks*, pp. 2445–2455 (2007)

6. Karnadi, F.K., Mo, Z.H., Lan, K.C.: Rapid Generation of Realistic Mobility Models for VANET. In: Wireless Communications and Networking Conference, pp. 2506–2511 (2007)
7. Naumov, V., Gross, T.R.: Connectivity-Aware Routing (CAR). In: Vehicular Ad Hoc Networks, the 26th IEEE International Conference on Computer Communications, pp. 1919–1927 (2007)
8. Zhao, J., Zhang, Y., Cao, G.: Data Pouring and Buffering on the Road: A New Data Dissemination Paradigm for Vehicular Ad Hoc Networks. IEEE Transactions on Vehicular Technology, 3266–3277 (2007)
9. Google Maps, <http://maps.google.com.tw/>
10. QualNet Network Simulator, <http://www.scalable-networks.com/>
11. SUMO, <http://sumo.sourceforge.net/>

Overlapping Community Detection with a Maximal Clique Enumeration Method in MapReduce

Yi-Jen Su, Wei-Lin Hsu, and Jian-Cheng Wun

Department of Computer Science and Information Engineering,
Shu-Te University, Kaohsiung City, Taiwan
iansu@stu.edu.tw

Abstract. Overlapping community detection is progressively becoming an important issue in social network analysis (SNA). Faced with massive amounts of information while simultaneously restricted by hardware specifications and computation time limits, it is difficult for clustering analysis to reflect the latest developments or changes in complex networks. To meet these demands, this research proposes a novel distributed computation method, which combines MapReduce, a distributed computation framework, and the TTT algorithm, to speed up the discovery of all maximal cliques in large-scale social networks. Then, overlapping community detection is implemented by the Clique Percolation Method (CPM) to incrementally merge adjacent cliques based on k -cliques with $k-1$ common nodes. Six groups of YouTube datasets (from 50K to 300K nodes with interval 50K) are adopted to evaluate clustering quality and execution time of the proposed method.

Keywords: Social Network Analysis, Overlapping Community Detection, MapReduce.

1 Introduction

Community detection refers to the identification of meaningful structures or concrete organizations in large-scale complex networks. This kind of techniques have been widely adopted in various research domains, including biology, sociology, communications, and so on. Social Network Analysis (SNA) presents the complex network as a Graph, G , formed by two components, nodes and edges. Each node n represents an actor, while each edge e stands for the explicit/implicit interactions between two nodes. Social Network Analysis [1] techniques detect social structures and special behaviors in $G(N,E)$. Users joining an interested user group on a social network will automatically form an explicit social group with the other members in it. Most implicit social groups embedded in complex social networks, however, can only be detected by interaction clues.

A community is defined as a cohesive subgroup with highly dense intra-group relationships and loose inter-group relationships. Community detection methods identify subgroups based on features of social networks. Such methods can be classified into

four categories: node-centric, group-centric, network-centric, and hierarchy-centric. Node-centric community detection is a traditional type of SNA methods, including clique, k-clique, k-core, and so on. Clique [1], for example, is a complete graph, in which each node owns direct links to all the other nodes within the same subgroup. In group-centric approaches, the density level of intra-group relationships has to be higher than a threshold. Network-centric approaches, such as block model approximation, measure the similarity of node pairs to divide the whole network into disjoint subgroups. Hierarchy-centric community detection distinguishes between divisive hierarchical clustering (top-down procession) and agglomerative hierarchical clustering (bottom-up procession).

These community detection methods can be further divided into two styles—disjoint and overlapping—depending on whether a node belongs to one group or multiple groups after community detection. Traditional community detection methods, for example, the GN algorithm [2], used to discover disjoint communities. However, on online social networks like Facebook, Twitter or Line, a user might simultaneously participate in several communities of interests while playing different roles in each of them, such as a family member, a friend, or a colleague. In such cases, disjoint community detection methods will misleadingly classify overlapped actors into different communities. The detection of overlapping communities [3] therefore becomes a crucial research issue.

In this study, the clique-finding process combines MapReduce [4] and the TTT algorithm [5] to speed up the identification of all maximal cliques in given large-scale complex networks. The MapReduce framework is adopted to implement distributed computing, which can retrieve relationship information about nodes and their neighbors in a given graph. The TTT algorithm, which is a maximal clique enumeration (MCE) method [7], modifies the BK algorithm [6] by adding the pivot nodes operation to prune recursive searches. Then, the overlapping community detection process is implemented by the Clique Percolation Method (CPM) [8] to incrementally merge adjacent cliques based on k-cliques with k-1 common nodes. This process is necessary because, if the clique requirement is too strict, the subgroup size will be too small to represent real life cases.

2 Literature Review

2.1 MapReduce

Google released the MapReduce framework in 2004. The system is implemented on Hadoop cloud platform developed by the Apache Foundation. MapReduce is a user-friendly framework that can be easily modified to meet distributed computing needs. First, the JobTracker of the master node partitions the assigned job into several tasks and then send them to the TaskTrackers of the data nodes to construct a cluster computing environment. The operations of MapReduce are divided into two parts: Map and Reduce. The Map operation parses the input data into key/value sets. Then the Reduce operation sorts Map results and generates outputs by merging same-key records.

2.2 Maximal Cliques

In SNA, most scholars agree that a high-quality subgrouping result has cohesive intra-group relationships and sparse inter-group relationships. Community detection by clique percolation will generate high-quality grouping results, but to detect all Maximal Cliques from a complex network is an NP-complete problem [9]. After the Maximal Clique Enumeration (MCE) process, for example, a given graph as shown in Fig. 1(a) is found to contain three cliques including $\{1,2,3,4\}$, $\{3,4,7\}$, and $\{5,6,7\}$, as shown in Fig. 1(b). The BK algorithm, proposed by Bron and Kerbosch in 1973, is a sequential DFS algorithm that enumerates all the maximal cliques of an undirected graph. Tomita *et al.* improved the BK algorithm with the pivot process to reduce the recursive times in DFS.

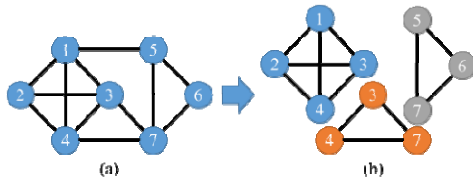


Fig. 1. An example of Maximal Clique Enumeration: (a) Given graph; (b) MCE results

2.3 Clique Percolation Method

When the clique mining method is adopted to find communities in complex networks, although all communities have the highest-density intra-group relationships, it is hard to get big subgroups. The clique constraints might be too strictly defined to respond to changes in real life. In essence, CPM discovers overlapping communities by incrementally merging adjacent k -cliques until none are detected anymore. In Fig 2(a), a 6-node graph, two cliques $\{1,2,3,4\}$ and $\{3,4,5\}$ are about to be merged by CPM with two adjacent 3-cliques $\{1,3,4\}$ and $\{3,4,5\}$ sharing two common nodes 3 and 4, as shown in Fig 2(b).

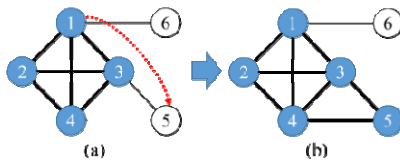


Fig. 2. Using 3-cliques in CPM: (a) Given graph; (b) CPM result

After CPM completes overlapping community detection, two communities $\{1,2,3,4,7\}$ and $\{5,6,7\}$ are found from a given undirected graph Fig 3(a), where node 7 simultaneously belongs to these two subgroups, as shown in Fig 3(b).

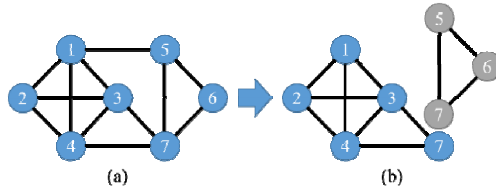


Fig. 3. Overlapping community detection: (a) Given graph; (b) Community detection result

2.4 Cluster Coefficient

Cluster Coefficient [10] is adopted to analyze the density level of relations between nodes within a group. A group with a larger number of intra-relations has more cohesion and a higher density value. In Eq. (1), N_i is the set of nodes, E_i is the set of edges, and K_i is the number of nodes in community C_i . The study uses the average cluster coefficient to evaluate the overall group quality. In Eq. 2, n represents the number of detected communities. The average cluster coefficient is computed by summing up all cluster coefficient values of each subgroup and then dividing the sum by the number of subgroups.

$$C_i = \frac{|\{e_{jk}: v_j, v_k \in N_i, e_{jk} \in E_i\}|}{\frac{1}{2}k_i(k_i-1)} \tag{1}$$

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i \tag{2}$$

Fig. 4 shows two groups $C_1=\{1,2,3,4\}$ and $C_2=\{3,5,6,7\}$. In group C_1 , the number of edges is 6 and k_i equals 4. In the other group C_2 , the number of edges is 5 and k_i is 4. After Eq. (1) is used to compute the value of cluster coefficient, the value of C_1 is 1 and C_2 is 5/6. The values of the average cluster coefficient is $(1+5/6)/2=11/12$.

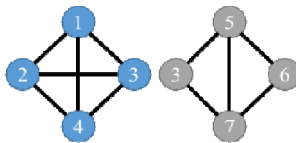


Fig. 4. Given graph for Cluster Coefficient calculation

3 Research Method

This study applies the distributed computing method MCE to overlapping community discovery as proposed by Bin Wu etc. in [11]. To prevent clique enumeration from generating cohesive subgroups of small sizes, the study proposes incrementally merge adjacent 3-cliques (subsets of any cliques having more than 2 member nodes) to complete the community detection. The clique-merging method, 3-clique CPM, is similar to combining two triangle graphs when both share two common nodes.

Following community detection, the whole operation can be divided into three steps: node connection retrieving, maximal cliques finding, and overlapping community discovering. The first step completes the relationship finding process in a two-level MapReduce operation that involves each node’s neighbors and each neighbor’s neighbors. Next, the maximal clique enumeration process uses the TTT algorithm to speed up the detection of all maximal cliques. Finally, the CPM is implemented by 3-clique merging to achieve overlapping community detection. To retrieve the information of each node’s connection in a given large-scale graph, a distributed computing method is developed in the MapReduce framework. The Mapper rearranges the input data in the same key-value format as the input of the Reducer, while the Reducer collects same-key records to represent the neighbor information of each node.

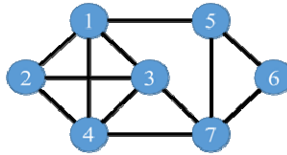


Fig. 5. Given undirected graph

The input of Level-1 MapReduce is the edge sets, as shown in Fig. 5. The Mapper collects all edges and transfers them to the key-value format. The node 1, for example, serving as the key, together with its four neighbors, nodes 2, 3, 4, and 5, become the key-value set {1-2,1-3,1-4,1-5}, as shown in Fig. 6. Then the Reducer puts the same-key records together, e.g. {1-2,3,4,5} to represent the neighbors of node 1.

Input	Map	Reduce
	key-value	
1-2	1 2	1 2,3,4,5 2 3,4 3 4,7 4 7 5 6,7 6 7
1-3	1 3	
1-4	1 4	
1-5	1 5	
2-3	2 3	
2-4	2 4	
3-4	3 4	
3-7	3 7	
4-7	4 7	
5-6	5 6	
5-7	5 7	
6-7	6 7	

Fig. 6. Level-1 MapReduce operation with a given graph Fig. 5

The operation of Algorithm 1 is divided into two parts: Map and Reduce. Using all edges of the given graph as the input, the Mapper not only converts the representation of all edges into the key-value format, but also arranges them in the ascending

order based on the key value. Removing duplicate edges can effectively speed up the Reducer operation. The output of the Reducer is an edge set with the same key, node k , as the neighbor list, $list(k)$.

Level-2 MapReduce uses a node's neighbor list to retrieve information about each neighbor's neighbor nodes. Take node 2 in Fig. 7 as an example. The expression $\langle 2, \langle 3, 4 \rangle \rangle$ means that node 2 has two neighbors, nodes 3 and 4. Then the expression $\langle 2, \langle 3, 4 \rangle \rangle$ is transferred to $3, \langle 2, 4 \rangle$ and $4, \langle 2, 3 \rangle$. The expression $3, \langle 2, 4 \rangle$, for example, represents that node 3 and node 2 are neighbors with the same relationship as that between node 4 and node 2. The Reducer collects all key-value records and sorts them with the value of k . The ordered key-value set is adopted in the process of MCE. To avoid generating duplicate cliques, the key-value sets with the values 2, 3, 5, and 6 are removed from the output candidates.

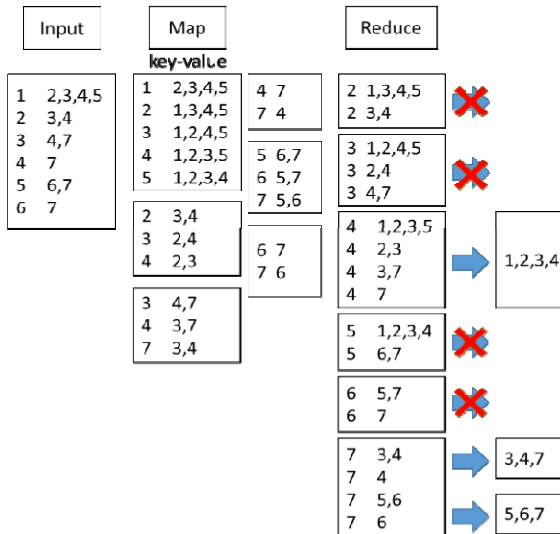
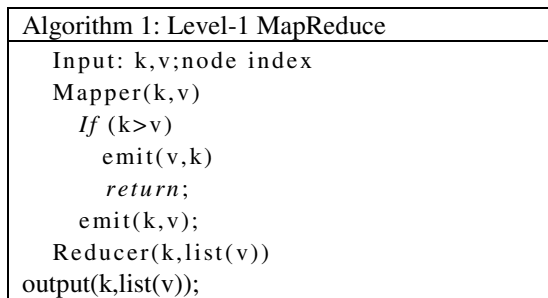


Fig. 7. Level-2 MapReduce operation with a given graph Fig. 5

The TTT algorithm is applied to enumerate all maximal cliques in the given graph. Fig. 8 shows an example of the TTT algorithm operation, where \square represents the node, u , with the biggest degree and \bullet represents the node with a direct link to node

u. In DFS, all ● nodes are not processed until the next level. The symbol Δ is adopted to stop the DFS path from generating duplicate cliques. The symbol ∇ signifies that the index number of the node is larger than that of the key node, and then the path will not be explored. For example, when the key node is 4 and the path 4-3-1-2 is traversed, because $7 > 4$, the DFS path 4-3-7 at the third level is stopped.

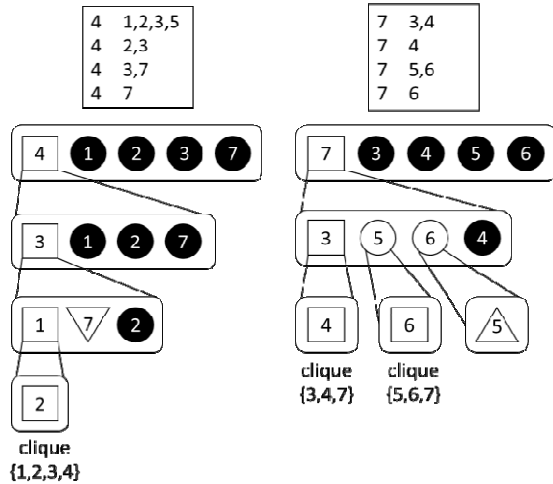


Fig. 8. Result of using TTT algorithm to implement MCE

In Algorithm 2, the Mapper uses the output of Level-1 MapReduce to exchange positions of the key node and the first neighbor node for providing information about each neighbor’s neighbors. The Reducer will pass the information to the TTT algorithm to enumerate all maximal cliques. The Reducer transfers $list(v)$, the relationship of a node and its neighbors, to G_{sub}^k , the relationship graph of node k . Both SUBG and CAND store the neighbor node list of node k .

Algorithm 2: Level-2 MapReduce

Input: k ; node index, $list(v)$, $\Gamma(k)$

Mapper($k, list(v)$)

$emit(k, list(v));$

for each $v' \in list(v)$

$emit(v', \{k\} \cup (list(v) - \{v'\}));$

Reducer($k, list(v)$)

$SUBG = \Gamma(k);$

$CAND = \Gamma(k);$

$Q = \{k\};$

$TTT(k, G_{sub}^k, SUBG, CAND, Q);$

4 Experiment Results

In this study, six sizes of YouTube user datasets (from 50K nodes to 300K nodes with interval 50K) are adopted to identify cohesive social structures. There are five observation criteria, including the number of subgroups, the average number of nodes in subgroups, the number of overlapping nodes, the average cluster coefficient, and execution time. In Fig. 9, as the dataset size grows, the number of detected subgroups also grows from 1015, 2143, ..., to 14040 nodes. The growth rate is faster when the dataset size is less than 200K nodes but slower when the dataset size is larger than 200K nodes. The number of overlapping nodes and the size of datasets are in direct proportion, as shown in Fig. 10.

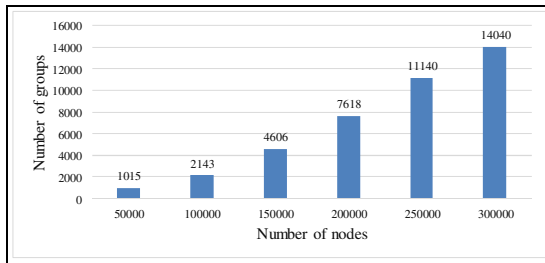


Fig. 9. Number of groups in community detection

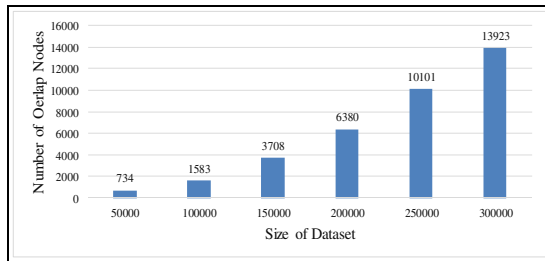


Fig. 10. Number of overlapping nodes in community detection

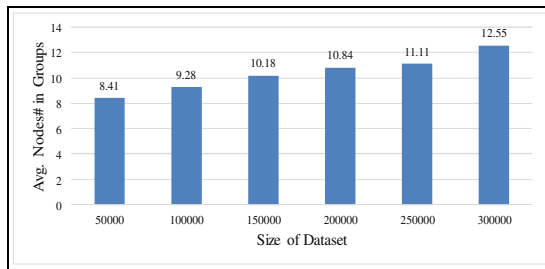


Fig. 11. Average number of nodes in groups

An interesting result, as shown in Fig. 11, is that, while the dataset sizes increase dramatically, the average number of nodes in groups grows very slowly. In terms of the overall quality of community detection, as shown in Fig. 12, the average cluster coefficient remains higher than 0.919. The results indicate high subgrouping quality and cohesion in all communities. Though CPM uses a looser constraint to detect communities, the result is still good enough for subsequent operations.

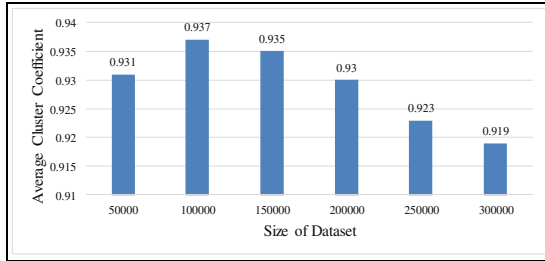


Fig. 12. Result of Average Cluster Coefficient

Using one machine to execute the CPM algorithm is time-consuming and requires large memory space. As the dataset sizes increase, the execution time increases sharply as well. In this study, there are three machines running on the clustering environment for community detection. The distributed computing architecture is apparently effective: the execution time grows steadily even when the size of dataset grows from 50K to 300K records, as shown in Fig 13.

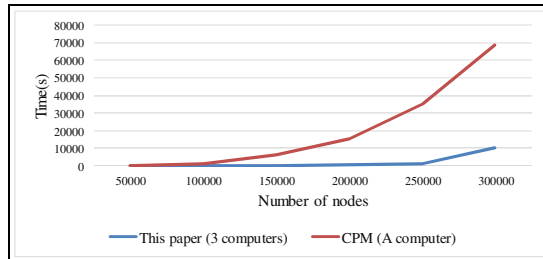


Fig. 13. Comparison of execution time

5 Conclusion

With so many social networks blooming, data sizes become considerably larger in problem domains than ever before. It is urgent to devise effective methods to speed up computation when solving big data problems. The distributed computation architecture is therefore developed in this study to detect overlapping communities in large-scale complex networks. Moreover, the study proposes using the MapReduce framework to facilitate the community discovery process. Using CPM to discover overlapping

communities also helps to avoid the problem of generating very small clique sizes. The experiment results sufficiently show that the proposed method is effective and efficient. Future research should continue to look for methods that further reduce computation time while attaining higher quality in social structure discovery.

References

1. Wasserman, S.: *Social network analysis: Methods and applications*. Cambridge University Press (1994)
2. Girvan, M., Newman, M.E.: Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* 99(12), 7821–7826 (2002)
3. Xie, J., Kelley, S., Szymanski, B.K.: Overlapping community detection in networks: The state-of-the-art and comparative study. *ACM Computing Surveys (CSUR)* 45(4), 43 (2013)
4. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. *Communications of the ACM* 51(1), 107–113 (2008)
5. Tomita, E., Tanaka, A., Takahashi, H.: The worst-case time complexity for generating all maximal cliques and computational experiments. *Theoretical Computer Science* 363(1), 28–42 (2006)
6. Bron, C., Kerbosch, J.: Algorithm 457: finding all cliques of an undirected graph. *Communications of the ACM* 16(9), 575–577 (1973)
7. Schmidt, M.C., Samatova, N.F., Thomas, K., Park, B.H.: A scalable, parallel algorithm for maximal clique enumeration. *Journal of Parallel and Distributed Computing* 69(4), 417–428 (2009)
8. Palla, G., Derényi, I., Vicsek, T.: The critical point of k-Clique percolation in the Erdős–Rényi graph. *Journal of Statistical Physics* 128(1-2), 219–227 (2007)
9. Michael, R.G., Johnson, D.S.: *Computers and Intractability: A guide to the theory of NP-completeness*. WH Freeman & Co., San Francisco (1979)
10. Stam, C.J., Jones, B.F., Nolte, G., Breakspear, M., Scheltens, P.: Small-world networks and functional connectivity in Alzheimer’s disease. *Cerebral Cortex* 17(1), 92–99 (2007)
11. Wu, B., Yang, S., Zhao, H., Wang, B.: A distributed algorithm to enumerate all maximal cliques in MapReduce. In: *Proceedings of the Fourth International Conference on Frontier of Computer Science and Technology, FCST 2009*, pp. 45–51 (2009)

Grey Analysis on Underwater Sensor Network of Penghu Set Net

Yih-Fuh Wang¹ and Chang-Ling Tsai²

¹ Department of Computer Science and Information Engineering & Graduate Institute of Electrical Engineering and Computer Science

² Graduate Institute of Electrical Engineering and Computer Science
National Penghu University of Science and Technology, Penghu, Taiwan
yfwang@npu.edu.tw

Abstract. The set-net enables the fishermen to collect the fish alive without harming them. Since the net is located near the shore, the set-net serves as a fishing bank because fishermen do not need to go far to fish. The fish harvest can be sold as soon as captured alive and be sold as value-added as branded fish. In this paper, we apply a small-scale underwater sensor networks for collecting data by the ratio of the information on tides and catches and use grey theory to do analysis. It reveals that the grey theoretical is possible to be helpful for analysis of the relationship between tides and catches in set net. Besides, simulation result shows that the grey correlation can be a better auxiliary to allow operators and consumers making decision when to send boat for catching.

Keywords: underwater sonar network, grey analysis, set net.

1 Introduction

The traditional set-net catch technology of Penghu was proud and invented some 30-40 years ago. The fish catch in the main trap net are able to swim around freely until fish caught. Set net fishery is a passive fishing methods and it is a larger trap of along coast. Every day, they catch fishes in the morning and afternoon in the same time, fishing season from September to next June. Although the catch time is fixed every day but it has a different tide, so the relationship between tides and catches are worth exploring. The past 30-40 years, Penghu set net fishing techniques have been proud of the job set-nets, fish traps will be directed to first, and then to the final cage, during which they stayed, waiting for boats to fishing nets.

For underwater explorations, classes of ad hoc networks, underwater ad-hoc networks (UANET) are used to explore large uninhibited oceans [1]. Unlike wireless links amongst land-based WSNs, each underwater acoustic link features large-latency and low bandwidth. Particularly underwater acoustic networks (UANs) [2] are an emerging technology for a number of oceanic applications, such as oceanographic data collection, offshore exploration, environmental monitoring and coastal surveillance. At first, a single point-to-point underwater high-data-rate link suffers

from a limited operational range. Secondly, an autonomous underwater vehicle (AUV) [3] plays the role in underwater missions to provide a collaborative communication with sensors of UWSN (see Fig. 1).

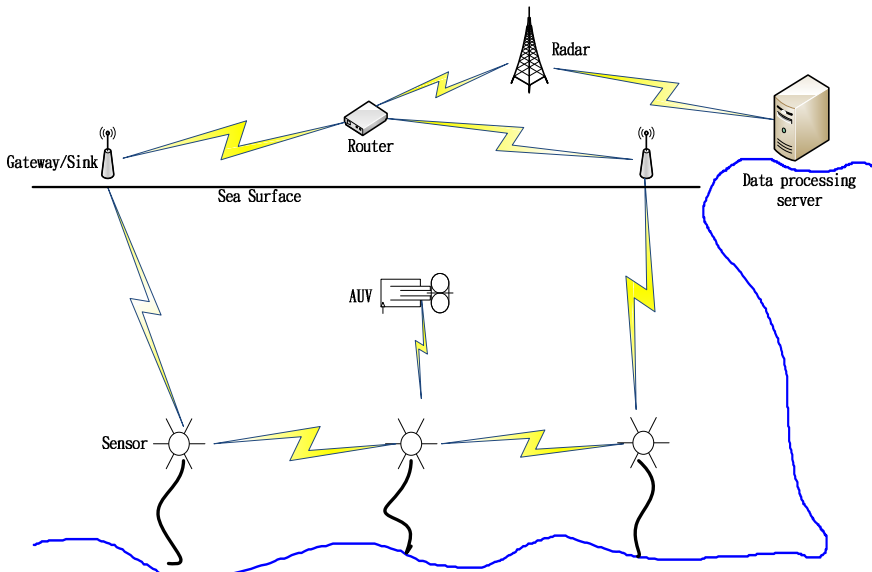


Fig. 1. AUV Underwater Monitoring in Large-scale UWSN

In UANs, nodes are usually fixed. No multiple mobile sensors are dispersing in UAN but UWSN is mobile and self-organized. Besides, for UANs, sensor localization is not desired since nodes are usually fixed. The sensor may be anchored in the sea floor or attached to a parking system. While enabling long ranges communication to a surface platform, the large-scale UWSN facilitates inter-communication between the AUVs to offer remote monitoring and control for the end user [3]. However, for this paper, the design of set-net underwater sensing system is based upon this concept of limited network deployment and hence we limited it in small-scale UAN. Meanwhile, due to wireless sensor are deployed for underwater circumstances and RF radio does not propagate for energy absorption, combined UAN with WSN on investigating of small-scale UWSN for set-net fishery are becoming essential.

In this work, this paper introduces one underwater sonar network [2] to deal with accurate information and real-time fishing in set nets and set up a wireless sensor network (WSN) [3] to help message transmission. For this reason, we develop a small-scale underwater sensor networks (UWSN) [3] in Penghu set net. From auxiliary sensing to monitor the UWSN, we can improve the efficiency of fishing operations. In addition, due to the efficiency, this method is reasonable and accurate rate to collect the information about tidal data and actual catches. Then, it will be able to assess the decision-making mechanism through grey theory [4]. Finally, we apply grey analysis to find out the rules in order to help operators and consumers on reducing boating fish about set net.

2 Underwater Monitor for Penghu Set Net

Generally a mobile UWSN is a scalable sensor network, which relies on localized sensing and coordinated networking among large numbers of low-cost sensors. In UWSNs, thousands of sensors are often deployed in hostile environments, battlefield, or security areas. Therefore, the sensor positions are random and unknown. In contrast, an existing UAN is a small-scale network that relies on data-collecting strategies like remote telemetry or assumes that communication is point-to-point. In remote telemetry, data are remotely collected by long-range signals. While in mobile UWSNs, sensor nodes are densely deployed in order to achieve better spatial coverage and thus a well monitor protocol is essential to avoid boating without

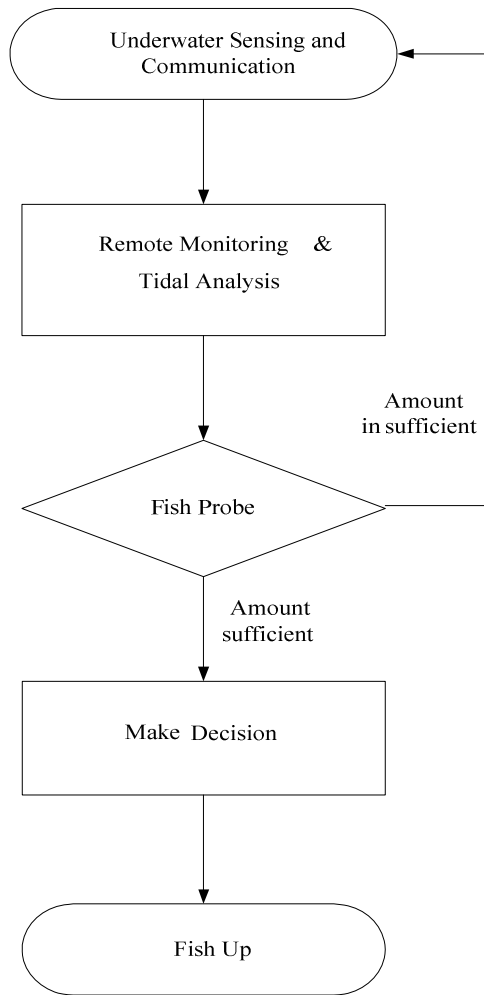


Fig. 2. Monitor Procedure on Set-net Fishing

enough harvest and improve the system throughput. Since the efficiency of this method can be unreasonably high for meeting the demands of high-precision, in our small-scale UWSN system, sensor nodes are usually sparsely distributed in local set-net and thus no serious multi-access technique is needed for consideration.

Each underwater wireless sensor nodes composing the UANs is real-time, so the fish-up judgment becomes a critical issue. Mainly our paper deals with the problem of accurately tracking the amount of fishes through combining sea-surface WSN and undersea UAN (WSN/UAN) to employ harvest sensing in set-net. Fig. 2 addresses the issues of estimating the tidal fish probing, improving judge efficiency by applying remote monitor and tidal information. We provide a wake-up fishing or fish up algorithm which increases decision making efficiency of probing the amount of set-net fishes through amount assessment system depicted in Fig. 2.

At beginning, we catch the signal of underwater sensing in UAN and transmit the information to the monitor by WSN. By tidal prediction assisting, we generate the judgment about fish-up or wait for fisherman. The flowchart in Fig. 2 represents the procedures of monitor algorithm for set net. Assisted from Fig. 1, in the first step, from the random sensor distribution, we perform the underwater sensing for fishes which are in overlap area of large range initial playground and leader-net in set-net. At this stage we consider only to put four sinks and if we determine whether the fishes is coming and its positions or directions in overlap area, the 1st step of localization algorithm end. In the second step, we apply the tidal information to assist the fish-up judgment by WSN data transmission and received by monitor. In the 3rd step, we calculate the amount of fishes by probing system. In the last step, we make a decision about fish-up alarm. If the amount of fishes is not enough, it will pass alarm. Otherwise, if the amount is enough, system will sent an alarm signal to fisherman to catch the fishes.

3 Grey Decision for Alarming Catches

In the proposed work consisting of variable number of sensors that are deployed to perform collaborative monitoring tasks over a given area of set-net in Penghu, to assess the aqueous environment, its role and function for the need of small-scale short term UAN and distributed information collection networks in WSN for periodic oceanic monitoring. We have to evaluate the efficacy of the proposed scheme in terms of the precisely information processing, which is a measure of real set-net performance based on Penghu tidal information.

We collected set net catches in Penghu and corresponding to the prevailing tide for past five years. Then we collect the total catches data for each tide with the number of fishing tides. In accordance with the principle of grey decision analysis, the catches will be average and unified standardized in the different tides. Finally, we use the grey decision analysis, pointed out that the association during the tides and catches objectively. The grey decision information is simplify and excludes the unnecessary attributes. Grey system theory in solving small size of the sample and discrete data, then input the useful data [5]. It is different to traditional statistical methods require

large data sets which use multiple variables to handle with small data sets and the distribution is not Normal [5].

To calculated and average from the different characteristics that is called the grey correlation [6]. Through grey correlation analysis the best combination of parameters can be used as resolution [6]. In the grey relational analysis is necessary to standardize the original data, because if the data is too large that some of the factors will be ignored [6]. However, after generated decision rules based on grey relational analysis to propose grey decisions. Grey decision analysis is a new method and has high practical value, the decision is feasible. We depicted the numerical result in Fig. 3.

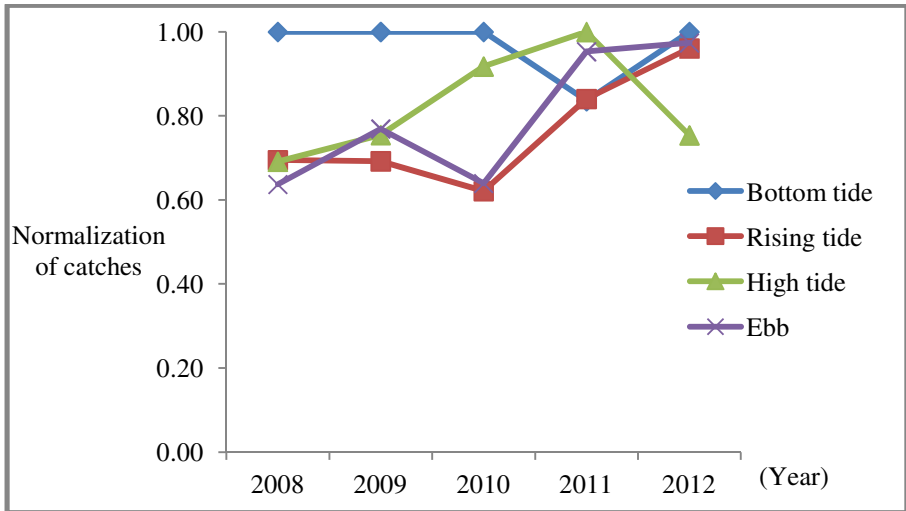


Fig. 3. Normalized grey value of catches vs. five years relations on different tides

Figure 3 indicates the relevance clearly of the catches and the tides during 2008 to 2012. Apart from 2011 that the bottom tide was poor performance, the rest four years are the most dazzling. Although the bottom tide was not good in 2011, it is insignificant on the other three tides. As the three tides leading, the difference distance was small and unstable on performance analysis. However, in these leading years of the bottom tide, the catches demonstrate significantly and stability more than the other three tides. Therefore, we can get information from the grey analysis and decision making and to make an effective judgment between the catches and the tides. This result tells us that catches is preferably on the bottom tide and it is the point for the best time to catch fishes.

4 Conclusion

This paper introduces a new class of underwater acoustic networks within a WSN communication for set-net monitor. It needs to be deployed for small-scale UWSN and face significant challenges in Penghu fishery. For set-net catches, the tide is very

important and inseparable. This numerical results point out how to determine the best time for fishing is feasible via the underwater sonar and wireless sensor networks. From the decision making purpose, grey analysis can help us to judge the relationship of tides and catches. Through the analysis of grey decisions, our approach can allow operators to reduce costs and consumers to get the best time on set-net fishing in Penghu.

References

1. Srinivas, S., Ranjitha, P., Ramya, R., Narendra, G.K.: Investigation of Oceanic Environment Using Large-Scale UWSN and UANETs. In: Proceedings of International Conference on Wireless Communications Networking and Mobile Computing, pp. 1–5 (2012)
2. Tomasi, B., Toso, G., Casari, P., Zorzi, M.: Impact of Time-Varying Underwater Acoustic Channels on the Performance of Routing Protocol. *IEEE Journal of Oceanic Engineering* 38(4), 772–784 (2013)
3. Yu, C.H., Lee, K.H., Choi, J.W., Seo, Y.B.: Distributed Single Target Tracking in Underwater Wireless Sensor Networks. In: Proceedings of SICE Annual Conference 2008, pp. 1351–1356 (2008)
4. Xie, M., Xiao, X.: Grey decision rules for interval MADA based on rough set theory. In: Proceedings of Grey Systems and Intelligent Services, pp. 866–869 (2011)
5. Lin, C.C., Lin, J.F., Yu, C.C., Lee, T.Q.: Proceedings of Machine Learning and Cybernetics, pp. 2898–2903 (2010)
6. Ranganathan, S., Senthilvelan, T., Gopalakannan, S.: Multiple Performance Optimization in Drilling of GFRP Composites Using Grey Analysis. In: Proceedings of Advances in Engineering, Science and Management, pp. 12–18 (2012)

A Research of Wireless Energy Collector for Increasing the Power of Rechargeable Device

Chuen-Ching Wang and Chi-Hung Wei

Dept. of Information Technology, Kuo Yuan University, Kaohsiung, Taiwan
t90261@cc.kyu.edu.tw
urderrick@gmail.com

Abstract. This study aims to develop a simple, low-cost, and unpolluting radio energy collector, which can transform the radio in the air into DC (direct current) source to be stored in the super capacity so that the idea of energy conservation can be fulfilled. To verify, the practicality and the feasibility of this convertor, the researchers use the vertical build-up antenna and the rectangular patch antenna to collect the EM wave (electromagnetic wave) emitted from the GSM base station and transform it into the DC source to supply the RFID Active Tag with power. Besides, a 433 MHz, 5W transmitter is used to imitate the device of high –amplitude wave in the indoor or underground parking lots to verify the feasibility of this system. The results indicate that the power, collected in the place about 50cm away from the device of the high –amplitude wave, can replace the battery of the RFID Active Tag.

Keywords: Energy Collection, RFID, radio power transmission, base-station.

1 Introduction

For the convenience in life , most people hope to substitute the wireless communication for the wired communication[1],[2]. In this case, more and more GSM base stations, repeater stations, broadcast station, and the wireless transmitting stations for the personal-use wireless devices are required to transmit the EM wave in all kinds of frequency to meet everyone’s needs[3].

Intercepting the energy in the air is the key point of the space activity and is one of the feasible programs for energy displacement as well. Today several programs have been proposed. The first one is UPS (Utility Power Satellite) in space [4] in which a device makes use of the sun, the nuclear energy, or other techniques to produce the energy, then the energy is transformed into microwave or laser beam, and finally the microwave or the beam are conveyed to the receiving station on the ground and further converted into the power to use. The second one is the Rectenna [5] which combines the rectifier with the antenna while the third one is W.C. Brown’s Powered Helicopter [6]. The forth one is the SPS (Solar Powered Satellite)[7] while the fifth is the SHARP (Stationary High Altitude Relay Platform)[8]. These programs are expected to provide another choice of power source by the WPT (Wireless Power Transmission). However, the programs above still have two problems to solve. One is

reduce the influence of the wireless power transmission of heavy power upon the earth environment, animals and plants, and people's bodies. The other is to promote the efficiency of transforming. If these two problems are solved, the programs proposed above can come true and replace the traditional polluting power source policy.

The air is filled with a variety of EM waves, such as the high-power radar wave used by the military, the heavy-power EM wave from broadcast stations, the EM wave from GSM base stations, and the EM wave from personal mobile communication devices. The energy is transmitted into the air all the time, but the quantity of the energy can be used is far less than the total quantity of transmitted energy. That is, a great quantity of earth resources are wasted in the air every day. Therefore, this study, on account of the concept that many a little makes a mickle, aims to develop an unpolluting, low-cost circuit of radio energy transformation to collect the energy of the EM wave in the air, transform it into the DC source to reuse, and replace the light-current systems, such as the power source of the mobile electronic products. The researchers have verified that this circuit can be applied to the Active Tag of the active RFID. That is to say, in some conditions, the active tag of the RFID does not need to use the battery.

The other sections of this study are displayed as follows. Section 2 displays the explanation of the structure of the system. Section 3 explains the designing theory, methods, and the practical approximate value of every unit circuit in the structure. Section 4 shows the results of the experiment while Section 5 comes to the conclusion.

2 The Framework of the System

Figure 1 shows the structure of the energy-receiving system. As shown in Figure 1, the down link frequency of the cell phone GSM base station is 1.805~1.880 GHz. In this paper, the down link frequency of the GSM base station, measured by the Spectrum analyzer, turns out to be 1.82GHz, so the resonant central frequency of the antenna is set at 1.82 GHz. In addition, this paper adopts the vertical antenna with quarter wave length and the rectangular patch antenna. The resonant circuit mainly uses the parallel resonant circuit. In order to regulate the resonant frequency easily, the capacitor of the parallel resonant circuit adopts the adjustable capacitor. Besides, the inductance is encircled by the insulating enamel-covered wire. The rectification circuit mainly uses the high-frequency diode. The electric capacity is used to store the DC source while the WAVETREND active tag is used as the testing machine.

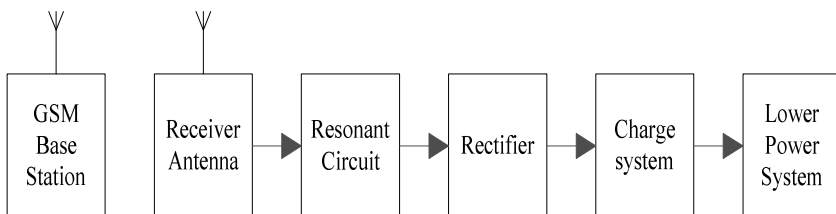


Fig. 1. The structure of the system

3 Theory and Design

A. The Resonant Circuit

When the circuit causes the resonance, the frequency which the resonance is correspondent to is called the resonant frequency and is shown as f_r or f_0 . When the resonant frequency occurs, the energy of the electrical capacitor is the same as that of the inductor, that is, $Q_L=Q_C$. When the energy is released from one reactance element, another reactance element will receive the same amount of energy. And these two reactance elements will produce energy pulsation as shown in Figure 2. Therefore, when the ideal (i.e. pure electric capacity and pure inductance) resonant circuit reaches the resonant state, they do not need extra reactive power and can keep resonant by themselves. This paper applies this effect to produce the highest output voltage to promote the rectification efficiency of the diode. However, in fact, there is equivalent resistance existing in the real inductor. And the equivalent resistance, working with the reactance element, produces appropriate damping effect and control the resonance curve.

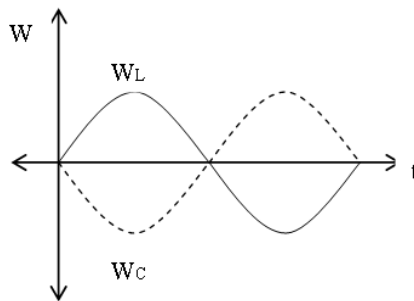


Fig. 2. The Energy Pulsation Drawing of the L and C Formation while Resonating

When the circuit reaches the resonant state, the input impedance of the parallel resonant circuit reaches the highest, and the terminal voltage of the two ends of the circuit also reaches the highest. In this case, the rectifying efficiency of Schottky barrier diode can be promoted. It is possible that the parallel resonant circuit produces resonance in other states of frequency or that the circuit cannot produce resonance. However, such conditions still can be avoided if the adjustment of the capacitance and the inductance value in wiring is well controlled. When the circuit reaches the output impedance and the output terminal voltage of the series resonant circuit are at its minimum value and will reduce the rectifying efficiency of Schottky barrier diode. Therefore, to promote the rectifying efficiency of Schottky barrier diode, this paper adopts the parallel resonant circuit. Figure 3 (a) shows the circuit of parallel resonance, and Figure 3 (b) shows the real equivalent circuit of the parallel resonant circuit. When the resonant circuit produces resonance, its resonant frequency is f_0 as shown in Eq. (1).

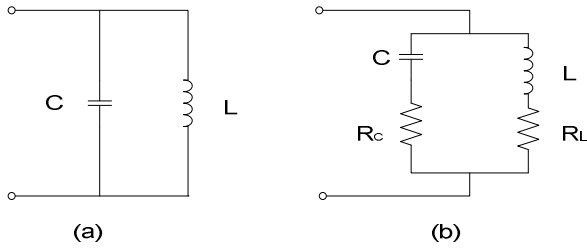


Fig. 3. (a) parallel resonant circuit (b) equivalent circuit of the parallel resonant circuit

$$f_0 = \frac{1}{2\pi\sqrt{LC}} \sqrt{\frac{L - CR_L^2}{L - CR_C^2}} \tag{1}$$

- f_0 : the frequency of parallel resonance ,Hz
- L: the inductance of parallel resonance ,H
- C :the capacitance of parallel resonance ,F
- R_L : the resistance of the inductance branch , Ω
- R_C =the resistance of the capacitance branch , Ω

According to Formula (1), some vital characteristics of the parallel resonant circuit can be conducted and are stated as follows.

1) When the resistance (R_L, R_C) of the two branch circuits is very little, that is, $R_L \rightarrow 0, R_C \rightarrow 0$, the parallel resonant frequency is

$$f_0 = \frac{1}{2\pi\sqrt{LC}} \tag{2}$$

Generally speaking, when $X_L > 10R_L$ and $R_C > 10R_C$, Eq. (2) can be used.

2) According to Eq. (1), when the resonant frequency f_0 is changed, the L or C value can be adjusted and the value of R_L or R_C also can be changed. However, because of the resistance with the characteristic of dissipation energy, it is almost impossible used in the real resonant circuit.

3) In the real circuit, because the value of electric resistance of the coil is larger, it cannot be omitted. Therefore, in the condition of $R_C \rightarrow 0$, the resonant frequency f_0 of the parallel resonant circuit is

$$f_0 = \frac{1}{2\pi\sqrt{LC}} \sqrt{1 - \frac{CR_L^2}{L}} < \frac{1}{2\pi\sqrt{LC}} \tag{3}$$

4) The essential condition for parallel resonance is $Q_L = Q_C$, and then $X_L = X_C$. But in Eq. 1, if $L - CR_L^2$ becomes $L < CR_L^2$ or $L - CR_C^2$ becomes $L < CR_C^2$, the result, by way of extraction of a root in mathematics, will become an imaginary number. But in physics, it is impossible to have a frequency signal which is a imaginary number. Therefore, in the parallel resonant circuit, if $L < CR_L^2$ or $L < CR_C^2$ occurs, the resonant frequency does not necessarily occur in the parallel resonant circuit; that is, the circuit does not result in resonance.

B. Schottky Barrier Diode

Schottky barrier diode [9] consists of metal and N-type semi-conductor, so the current carrier is mainly the electrons. Because the electrons move faster than the holes, its speed of switching is much faster than the common diode. Therefore, Schottky barrier diode can be used as a high-frequency rectification diode. According to the analysis of the equivalent circuit of the diode, the input impedance of the diode changes with the input power. Therefore, when the input power becomes different, the efficiency of rectification will change as well. In the same input power, the constant change of the resistance of the load will lead to the change of efficiency. Therefore, in a variable environment, the efficiency of the diode can be just a general value, which is commonly about 360Ω. In the practical performance, for the sake of the match of impedance, the circuit can be like the model of the equivalent circuit of the simplified diode as shown in Figure 4.

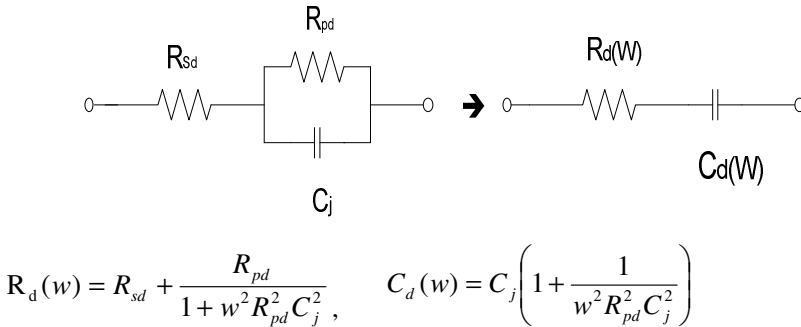


Fig. 4. The Equivalent Circuit of the Diode and the Drawing of the Simplified Equivalent Circuit

4 Experimental Results

Figure 5 shows that the HP Spectrum analyzer is used to test the down-link frequency of the GSM base station. As shown in Figure 5, the down-link frequency of the GSM base station is 1.82GHz, so the resonant frequency of the circuit in this study is set at 1.82 GHz. When 1.82 GHz is put into the formula $C=f\lambda$, we can get $\lambda=0.1648m$. Then, the wave length multiplied by the shortened rate 0.95 is $\lambda=15.66cm$. The wave length is marked as $1/2\lambda$, $1/4\lambda$, and $1/4\lambda$. As shown in Figure 6, the second $1/4\lambda$ is folded to be a virtual ground plane. This kind of accumulative antenna uses $1/2\lambda$ wave length in the upper part to from the resonant current and them strings and adds the resonant current of the $1/4\lambda$ wave length in the lower half part. Therefore, if the resonance in the upper and lower parts can be tuned well, its antenna gain compared with the gain of the antenna with $1/4\lambda$ wave length will increase 2-3dB.

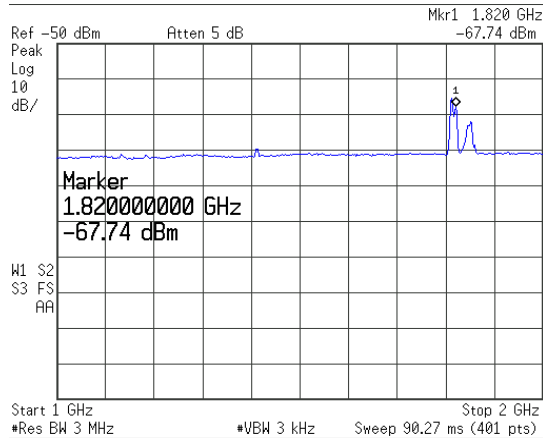


Fig. 5. The Down-Link Frequency of the GSM Base Station

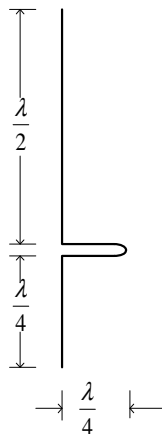


Fig. 6. The Accumulative Antenna

The vertical antenna with a quarter of the wave length is designed in the frequency 433MHz and receives the change-over circuit. In the indoor test, the 5W transmitter is used. The relation between its charging current for testing and the distance is shown in Figure 7. In the outdoor part, the transmitted power is 100W. The relation between its charging current for testing and the distance is shown in Figure 8. The result of the indoor test indicates that the charging current which is got at the location of 50cm can replace the battery of the active tag. The result of the outdoor test indicates that the charging current which is got at the location of 6m can replace the battery of the active tag.

According to Figure 7 and Figure 8, it is clear that the further the distance is, the less energy it can receive (Because the magnetic field intensity and the square of the distance are in inverse ratio). However, the EM wave in the air occurs 24 hours a day and many a little makes a little, so the collected energy is able to be used.

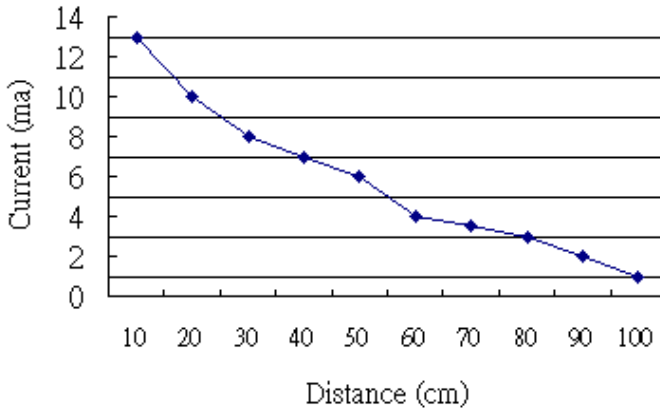


Fig. 7. The Charging Current and the Distance (5W@RL=261Ω)

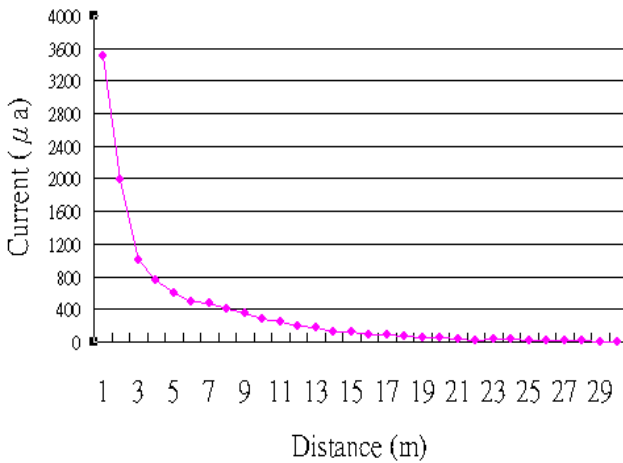


Fig. 8. The Charging Current and the Distance (100W@RL=261Ω)

5 Conclusions

This paper demonstrates that the radio EM wave in the air can be intercepted and be transformed into the DC power source to replace the battery of the system in need of weak power. We use the accumulative antenna 35m away from the GSM base station to store the electric current of $65 \mu A$ in the super capacitance. In the same condition, by way of the test of the rectangular patch antenna, we get the electric current $70 \mu A$. Besides, the transmitters of 433MHz, 5W and 100W simulate the electric wave in the air indoor and outdoor. The experiment demonstrates that the system in this paper can replace the battery of the active tag, and collecting the micro-energy by way of accumulation can promote the efficiency of conversion.

Acknowledgments. The Authors would like to thank the financial support from National Science Council and National Communication Council (NSC 102-2221-E-224-017-)in Taiwan.

References

1. Akyildiz, F., et al.: A survey on sensor networks. *IEEE Commun. Mag.* 40(8), 102–114 (2002)
2. Chin, D.: Nanoelectronics for the Ubiquitous Information Society. In: *IEEE ISSCC Dig. Tech. Papers*, pp. 22–26 (2005)
3. Rabaey, J.M., et al.: PicoRadio Supports ad Hoc Ultra-lowpower Wireless Networking. *Computer* 33(7), 42–48 (2000)
4. Solar Power Satellite Program Rev. DOE/NASA Satellite Power System Concept Develop. Evaluation Program. In: *Final Proc. Conf.* 800491 (1980)
5. Park, Y.-H., Youn, D.-G., Kim, K.-H., Rhee, Y.-C.: A Study on the Analysis of Rectenna Efficiency for Wireless Power Transmission. In: *Proceedings of the IEEE Region 10 Conference*, pp. 1423–1426 (1999)
6. Brown, W.C.: Experiments Involving a Microwave Beam to Power and Position a Helicopter. *IEEE Trans. Aerosp. Electron. Syst.* 5(5), 692–702 (1969)
7. Brown, W.C.: Solar Power Satellite Program Rev. DOE/NASA Satellite Power System Concept Develop. Evaluation Program. In: *Final Proc. Conf.* 800491 (1980)
8. Schlesak, J., Alden, A., Ohno, T.: A Microwave Powered High Altitude Platform. In: *IEEE MTT-S Int. Microwave Symp. Dig.* (1988)
9. ISS106 Silicon Schottky Barrier Diode for Various Detector, High Speed Switching. HITACHI

How to Determine the Best Indexes of Industry Website by FANP Approach

Chih-Chao Chung¹, Hsiu-Chu Huang², Huei-Yin Tsai³, and Shi-Jer Lou^{4*}

¹ Graduate Institute of Engineering Science and Technology, National Kaohsiung First University of Science and Technology, Taiwan

² Department of Leisure and Tourism Management, Shu-Te University, Taiwan

³ Department of Industrial Technology Education,
National Kaohsiung Normal University, Taiwan

⁴ Graduate Institute of Technological & Vocational Education,
National Pingtung University of Science & Technology, Taiwan
9915916@gmail.com

Abstract. This study is aimed to evaluate and select design indicators of enterprise web portal through FANP (Fuzzy Analytic Network Process) and establish FANP-based assessment model of design indicators of enterprise web portal. First, we discuss the relevant literature, and analyze demand assessment of web portal when enterprises introduce e-business. The important design indicators of enterprise web portal include five dimensions, totaling 15 design indicators. Based on FANP expert questionnaire analysis results, uncertainty and fuzziness of the expressions or decisions of experts are eliminated to effectively reflect views of the experts. The findings showed among the five web portal design dimensions, the dimension of two-way communication is the most important, followed by distribution, trading mechanism, webpage design and advertisement and promotion. In addition, priority ordering of the overall design indicators is obtained for analyzing the management implications, and the suggestions are provided.

Keywords: website, website design, indicator evaluation, FANP, fuzzy.

1 Introduction

With advancement of Internet, the transaction agreement can be concluded without talking face to face in changing market trading mode and buying and selling of any items. Medium and small enterprises should enhance their constitution and actively introduce E-business, so as to integrate them with the world. Besides, effective integration of industry resources and establishment of e-business functions can improve service for customers and lift company image. Then, the companies can expand market and adjust the marketing strategies to maintain their competitive advantages [1].

* Corresponding author.

Analysis of fortune global 500 homepages by many scholars such as Liu, Arnett [2] showed in terms of the use percent of each item in main contents the top three are product introduction/service (93.2%), company profile and information (86.1%), and interaction and feedback (79.3%). According to BizRate.com of United States, 67% of e-business consumers visit the shopping website via the network, and 33% find the shopping websites through other channels. Thus, network business performance is regarded as one part of the enterprises performance. Rao Kowtha and Whai Ip Choon [3] examined the correlations between previous core competence, scale, network development time, establishment time, competitive intensity and commitment of the companies. They suggested that these factors may affect development and performance of e-commerce websites.

If the traditional business methods are implemented to the network without change, new opportunities cannot be created. New business philosophy and model, and offering high added value service and Internet surfing experience for customers pose new challenges to enterprises in introduction of e-business. Huizingh [4] proposed important characteristics and design of websites; Ghose and Dou [5] indicated the higher the interactivity is, the higher appeal of the websites is. Thus, how to design characteristic web portal in response to the industry features is the problem medium and small enterprises will have to face in e-business introduction.

From the above, in consideration of function design of the enterprise web portal, this study used network hierarchy analysis method and fuzzy theory to establish FANP-based assessment mechanism of web portal indicators, and provide reference for design of enterprise web portals. The objectives of this study are as follows:

- 1) Discuss web portal design indicators.
- 2) Evaluate priority ordering of web portal design indicators.
- 3) Analyze web portal design management implications.

2 Literature Review

According to the research objectives, this study discusses and analyzes the web portal design and FANP literature, and they are described as follows.

2.1 Web Portal Design

Murtaza [6] defined web portal as an application tool for enterprises to access internal and external information, and it provides one gateway for users to obtain required information and make accurate decision. Angehrn [7] suggested four main spaces related to business in ICDT, (1) Virtual Information Space (VIS); (2) Virtual Communication Space (VCS); (3) Virtual Distribution Space (VDS) and Virtual Transaction Space (VTS). It is intended to guide enterprises to establish one set of strategy and procedure and assist them in design of commercial activities or develop and research new products and service. This can effectively increase liquidity and completeness of data, reduce consumer search and shopping time and improve convenience and satisfaction [8]. Enterprises can cope with competition and impact

by Internet, seize new opportunities derived from Internet development and become leaders of the Internet trading.

Besides, the interactivity as proposed by Deighton [9] includes three meanings: It can take demand ability of net friends into consideration, collect and store responsive ability of individual net friends; retake demand ability of individual net friends into account in terms of the response. Seybold and Marshak [10] also indicated introduction of e-business strategy by enterprises can bring seven benefits for customers. (1) Improve customer loyalty; (2) increase profitability; (3) accelerate launch of new products; (4) achieve customer goals through optimal cost efficiency; (5) reduce trading costs; (6) reduce customer service time; (7) reduce customer service time.

From the above, this study used ICDT as main architecture, and website interaction was considered in the design. The important indicators of web portal design in introduction of e-business by enterprises are summarized. The five dimensions include: (1) webpage design; (2) promotion; (3) distribution service; (4) trading mechanism, and (5) two-way communication, and there are 15 design indicators, as shown Table 1. FANP is used as the analysis architecture to select priority of web portal design indicators. This can assist enterprises in establishment of one set of effective web portal with complete functions, and improve enterprise competitiveness.

Table 1. Enterprise web portal design indicators

Interaction function	Description
1. Webpage design	(1) Rich information Post the latest information and new activities, make users close to websites, and attract new members.
	(2) Easy to operate interface In link with relevant information, users will not be confused among the web pages and will not lose the interest in shopping due to complicated operation procedure.
	(3) Site Navigator The site navigation function can make users familiar with the web architecture. The site navigation can be displayed through web map.
2. Promotion	(1) Advertisement In order to make more people know the products, increase product exposure and popularity and attract more potential customers.
	(2) Lottery/ prize The shopping websites provide lottery/prize for visitors. During these activities, promote corporate image or product.
	(3)e-paper Send e-paper to users, and make them know new products and the price or activities, and make consumers know and have more choices.
3. Distribution	(1) Customer database Store basic information of members, such as birthday, telephone, identity and address in the database, and deliver goods, mailing list, and e-paper.
	(2) Shopping cart This function can make consumers know the shopping list, confirm the purchased goods and facilitate customer shopping.
	(3) Personalized promotion The website recommends suitable goods for each customer as per their different demand characteristics.
4. Trading mechanism	(1) Online supervision The website service personnel irregularly monitor online trading, provide necessary assistance and consultation for customers and express comments through e-mail.
	(2) Trading record inquiry The website discloses historical trading records, and list product sale ranking list as reference for customers.
	(3) Data transmission security Establish the security mechanism in the website to prevent disclosure of personal account data.
5. Two-way communication	(1) Authentication Before the website provides online ordering service, identity of the buyer must pass the authentication and then transaction can be made.
	(2) Product inquiry In the website, the products are classified, and product specification, style and functions are provided to facilitate inquiry and shopping for users.
	(3) Communication and service Buyer and seller can make communication through "Q&A" function, and make product choice clear and reduce time for buyers in shopping.

2.2 FANP(Fuzzy Analytic Network Process)

ANP as proposed by Saaty [11] is aimed to overcome possible relation of interdependence between attributes or dimensions in AHP framework and feedback effect problems [12, 13]. ANP development uses the network flow concept to improve independent limitation caused by hierarchy architecture [14].

Lee and Kim [15] employed ANP method in zero-order target planning model to select information system schemes. This method can reflect relation of interdependence of assessment criteria in the goal programming, and make information system planning more efficient. In terms of the similar assessment models, Karsak, Sozer [16] used ANP and objective planning method to solve production process allocation problems in quality function deployment architecture. Chung, Lee [17] further used matrix evaluation concept as suggested by Saaty and simple ANP method, and analyzed the optimal production combinations of the semiconductor industry in multi-input and multi-output process.

In application of ANP and Fuzzy set, the decisions and judgments of people have certain difference, and are fuzzy and inaccurate. Thus, fuzzy set concept and algorithm were introduced to ANP to assist human in solving uncertainty of perceived evaluation. Generally, ANP emphasizes how to transform linguistic variables into fuzzy variables, and uses pairwise comparison matrices to obtain weight vectors. It is aimed to know evaluation characteristics of decision makers and effectively select optimal strategy schemes or development plans [18-21].

3 Research Design

Based on the past literature, the web portal design indicator assessment model was established with FANP and the analysis flow is as follows.

Step 1: Establish assessment architecture

Based on the litterateur analysis results, the important indicators for industry web portal design include five dimensions, and 15 design indicators. Based on the triangular fuzzy number concept, the fuzzy semantic evaluation set used by decision makers in reasonable architecture of fuzzy set is assumed to be N_k , N_k =[average, somewhat important, important, very important and absolute important] ($k=1, 2, 3, \dots, 5$) so as to evaluate relative importance of dimensions and design indicators.

Step 2: establish pairwise comparison fuzzy matrices

Based on the fuzz analysis method, fuzzy numbers are used to evaluate relative importance of design indicators under the web portal design dimensions. The comparison assessment criteria are the same as AHP, and nine criteria are used for a series of pairwise comparison. Geometric average is used to summarize expert opinions. This can increase consistency and accuracy of factor judgment. As shown in Eqs. (1) and (2), M_{ij} is median value of fuzzy numbers, and minimum and maximum score of the respondents is lower limit L_{ij} and upper limit U_{ij} of triangular fuzzy numbers. In this way, all the expert opinions can be transformed into fuzzy numbers.

Thus, triangular fuzzy number a_i is used to establish fuzzy pairwise comparison matrix \tilde{A} between the design indicators, as shown in Eq. (3).

$$M_{ij} = \sqrt[n]{\prod_{k=1}^n B_{ijk}} \tag{1}$$

$$\tilde{a}_{ij} = (L_{ij}, M_{ij}, U_{ij}) \tag{2}$$

$$\tilde{A} = \begin{matrix} & \begin{matrix} C_1 & C_2 & \dots & C_i & \dots & C_k \end{matrix} \\ \begin{matrix} C_1 \\ C_2 \\ \vdots \\ C_i \\ \vdots \\ C_k \end{matrix} & \begin{bmatrix} 1 & \tilde{a}_{12} & \dots & \tilde{a}_{1i} & \dots & \tilde{a}_{1k} \\ \tilde{a}_{21} & 1 & \dots & \tilde{a}_{2i} & \dots & \tilde{a}_{2k} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \tilde{a}_{i1} & \tilde{a}_{i2} & \dots & 1 & \dots & \tilde{a}_{ik} \\ \vdots & \vdots & \dots & \vdots & \ddots & \vdots \\ \tilde{a}_{k1} & \tilde{a}_{k2} & \dots & \tilde{a}_{ki} & \dots & 1 \end{bmatrix} \end{matrix} \tag{3}$$

where, $\tilde{a}_{ik} = \frac{1}{\tilde{a}_{ki}}$

Step 3: Defuzzification

This study used the equation proposed by Liou and Wang [22] for defuzzification, as shown in Eq. (4). In the equation, α is risk appetite of decision makers, and can be regarded as steady change state of real environment. When the value is 0, change range of the environment uncertainty is maximum; when the value is greater, the decision environment is relatively stable, and decision variation is smaller. λ is the risk exposure of decision makers, and the decision makers can assign different risk values as per different conditions. When it is 0, the decision makers regard the decision is low risk, and when it is 1, the decision makers regard the decision is high risk.

$$D_{\alpha,\lambda}(a_{ij}) = [\lambda * f_{\alpha}(L_{ij}) + (1 - \lambda * f_{\alpha}(U_{ij}))], 0 \leq \alpha \leq 1, 0 \leq \lambda \leq 1$$

$$f_{\alpha}(L_{ij}) = (M_{ij} - L_{ij}) * \alpha + L_{ij} \tag{4}$$

where, $f_{\alpha}(U_{ij}) = U_{ij} - (U_{ij} - M_{ij}) * \alpha$

Step 4: calculate eigenvalue and eigenvector

After defuzzification, they are transformed into single value. The calculation concept is the same as ANP. For equation of eigenvalue and eigenvector, as expressed in Eq. (5). Each comparative matrix must undergo consistency check, as shown in Eqs. (6) and (7). Saaty [11] suggested that CR is optimal when it is smaller than 0.1 This reveals judgment in the questionnaires of expert has consistency.

$$D_{\alpha,\lambda}(A) \times W = \lambda_{\max} \times W \quad \text{where,} \quad \lambda_{\max} = \sum_{j=1}^n W_j / W_i \tag{5}$$

$$CI = \frac{\lambda_{\max} - n}{n - 1} \tag{6}$$

$$CR = \frac{CI}{RI} \tag{7}$$

Step 5: Establish supermatrix

In supermatrix, all the groups and the included factors are prioritized in the matrix. If there is no dependency between criteria or factors, the value is 0 in the supermatrix, and namely the supermatrix can show dependency and relative importance.

Step 6: Calculate weights

The weight calculation includes three matrices, non-weighted supermatrix, weighted supermatrix and limiting matrix. The original supermatrix, i.e. non-weighted supermatrix, row vectors of normalized supermatrix, and sum of each row vector is 1. In this case, the supermatrix is weighted supermatrix. The weighed supermatrix is multiplied by $2k+1$, as shown in Eq. (8). At last, the convergent limit value can be obtained. This is limiting supermarix, and weight of the criteria and the factors can be obtained.

$$W_{sp} = \lim_{k \rightarrow \infty} W^{2k+1} \quad (8)$$

4 Research Results

This study designs two types of questionnaires: in-depth interview questionnaires of enterprise web portal design indicators, and FANP questionnaires of network design indicators. In in-depth interview questionnaires, the questionnaire design was based on website design literature. Experts were invited to determine which website design strategies can improve corporate overall efficiency when the network marketing is introduced. In FANP-based questionnaires, this study summarized the in-depth interview results and established hierarchical structure of design dimensions of enterprise web portal. The questionnaire design contains two parts, including: (1) fuzzy pairwise comparison questionnaires for web portal design dimensions; (2) fuzzy pairwise comparison questionnaires of relation of interdependence. The scope of investigation of the web portal design indicators is medium and small enterprises. The experts have rich experience and knowledge and were the respondents, as shown in Table 2. This study used “Years of experience” and “Specialty” as two conditions to select samples. The threshold value of “Years of experience” is ten years and “Specialty” is related to website design research, and we selected experts who satisfy the two conditions.

Table 2. Data sheet of interviewed experts

Expert	Years of experience	Specialty	Job Title
A	12	Multimedia system design/Human-Computer Interaction	Professor
B	13	e-commerce/data processing	Professor
C	10	Information system analysis and implementation	Associate Professor
D	12	Development of digital marketing system	Professor
E	10	Website design /CAD	Manager
F	12	Network marketing/customer management	General Manager
G	11	Website design/virtual reality	Manager

4.1 Establishment of Assessment Architecture

This study had five dimensions of web portal design: “webpage design”, “advertisement”, “distribution service”, “trading mechanism” and “two-way communication”, and 15 design indicators, as shown in Figure 1.

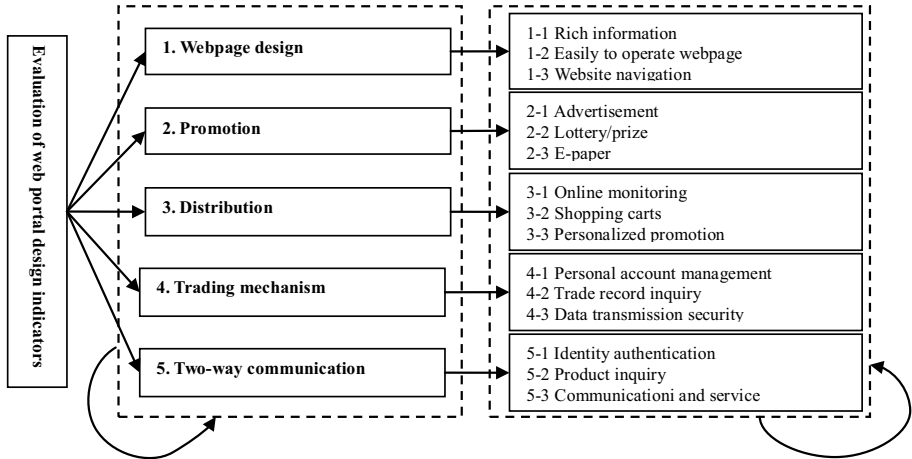


Fig. 1. Evaluation model of web portal design indicators

4.2 Establishment of Fuzzy Pairwise Comparison Matrices

By interviewing the experts, we conducted assessment for the importance of the web portal design dimensions and indicators, and integrated fuzzy pairwise comparison matrices of seven experts through calculation of fuzzy numbers integration operation. The evaluation results of importance of the design indicators under different website design dimensions are obtained, and illustrated by the five dimensions, as shown in Table 3.

Table 3. Fuzzy pairwise comparison matrices of experts under website design dimensions

Dimension	1. webpage design	2. advertisement	3. distribution service	4. trading mechanism	5. two-way communication
1.	(1,1,1)	(0.33,1.38,4)	(0.2,0.49,4)	(0.17,2.58,7)	(0.14,0.44,3)
2.	-	(1,1,1)	(0.3,0.44,4)	(0.14,0.45,5)	(0.11,0.49,4)
3.	-	-	(1,1,1)	(0.21,1.68,7)	(0.12,0.44,3)
4.	-	-	-	(1,1,1)	(0.09,1.28,5)
5.	-	-	-	-	(1,1,1)

4.3 Defuzzification

Before defuzzification, risk appetite of expert decision makers is set to 0.7, and risk exposure is set to 0.9. Next, Eq. (4) is used for defuzzification to obtain defuzzified matrices of each design indicator under different website design dimensions. The comparison matrices of the five dimensions are used for illustration, as shown in Table 4. Next, all the defuzzified matrices are used in evaluation of pairwise

comparison matrices and weight vectors to calculate eigenvalue and eigenvector. Hence, weighted value of each matrix is obtained, and consistency test is conducted. CR of the comparison matrices of the five dimensions is 0.077, and smaller than 0.1. This indicates that evaluation results of expert questionnaires have consistency.

Table 4. Defuzzified pairwise comparison matrices of experts under website design dimensions

Dimension	1.	2.	3.	4.	5.	Weight
1. webpage design	1	1.175	0.517	2.062	0.436	0.175
2. advertisement	-	1	0.509	0.503	0.493	0.116
3. distribution service	-	-	1	1.443	0.430	0.220
4. trading mechanism	-	-	-	1	1.070	0.187
5. two-way communication	-	-	-	-	1	0.302

CR=0.077<0.1

4.4 Establishment of Supermatrices

The dimensions and design indicators are prioritized in the matrices. Based on dependency between the dimensions or design indicators, the relevant weights are filled in to form original supermatrix, namely non-weighted supermatrix. In A1, non-weighted supermatrix, row vectors of normalized supermatrix to make summation of each row vector be equal to 1, and this supermatrix is weighted supermatrix. In A2, Eq. (8) is used to obtain converged limiting value, and this is limiting supermatrix. In A3, all the design weighted values are obtained after summarization, as shown in Table 5. The description is given in the next section.

Table 5. Summary table for weights of enterprise web portal design indicators

Dimension of website design	Weight	Design indicators	Weight	Priority
1. webpage design	0.175	1-1 Rich information	0.047	9
		1-2 Easy to operate interface	0.026	11
		1-3 Site Navigator	0.009	13
2. advertisement	0.116	2-1 Advertisement	0.081	5
		2-2 Lottery/ prize	0.009	14
		2-3 e-paper	0.042	10
3. distribution service	0.220	3-1 Customer database	0.050	7
		3-2 Shopping cart	0.068	6
		3-3 Personalized promotion	0.082	4
4. trading mechanism	0.187	4-1 Online supervision	0.020	12
		4-2 Trading record inquiry	0.008	15
		4-3 Data transmission security	0.049	8
5. two-way communication	0.302	5-1 Authentication	0.176	2
		5-2 Product inquiry	0.151	3
		5-3 Communication and service	0.182	1

4.5 Analysis on Weight Priority

The priority of the web portable design indicators in introduction of e-business can be determined by priority ordering. The overall situations and dimensions are analyzed and described as follows.

1) Overall analysis: first, the expert question analysis results under the website design dimensions show (5) two-way communication is the most important (0.302), followed by (3) distribution service (0.220), (4) trading mechanism (0.187), (1) webpage design (0.175), and (2) advertisement (0.116). This indicates most of experts considered enterprise web portal design must be focused on two-way communication design followed by distribution. Next, in terms of overall performance of the design indicators, the top five indicators are 5-3 communication and service (0.182), 5-1 identity authentication (0.176), 5-2 product inquiry (0.151), 3-3 personalized marketing (0.082), and 2-1 advertisement (0.081). This indicates most of experts considered customer communication and service, customer identity authentication and customized product inquiry and personalized sales promotion are more important among the web portal design indicators. Besides, advertisement of company products should be also emphasized.

2) In analysis of each dimension, under the (1) website design dimension, 1-1 rich information (0.047) is the most important. Under (2) promotion dimension, priority is given to 2-1 advertisement (0.081). Under (3) distribution dimension, attention must be paid to fulfillment of 3-3 personalized sales promotion (0.082). Under (4) trading mechanism, the focus is on improvement of 4-3 data transmission security (0.049). Under (5) two-way communication dimension, 5-3 communication and service (0.812) is the most important.

5 Conclusions and Management Implications

In order to understand experts' views on importance of enterprise web portal design indicators in introduction of e-business, this study used FANP to obtain relative weights of the 15 design indicators to know their priority.

The findings show that enterprise web portal design must achieve objective of two-way communication between enterprises and customers. In terms of the overall analysis, the top five indicators are communication and service, identity authentication, product inquiry, personalized promotion and advertisement. Based on customer-oriented standpoint, enterprises can provide better service, safe shopping environment, and thoughtful shopping functions to increase shopping convenience of customers and added value.

Regarding the management implications, the proposed FANP assessment model of enterprise web portal design indicators can be used before enterprises introduce e-business. Based on the industry attributes, enterprises can select suitable design indicators to design web portal with company characteristics. The e-business enterprises can establish a project team, and the team can use the FANP assessment model to review advantages and disadvantages of existing web portal design, and provide reference for further improvement. Enterprises can make e-business marketing strategy planning with reference to the research results. It is necessary to provide complete customer service and product information, and detailed product introduction information can reduce doubt of buyers and increase purchase willingness of consumers. Although Internet is highly developed, in the virtual

transaction behavior, buyers still concern the transactions. Thus, safe shopping environment is also important. Before successful transactions, safe payment methods and inquiry service network shopping progress shall be provided to ensure smooth transactions and bring out trade agreement between buyers and sellers. In addition, detailed customer transaction records, and personalized service and advertisement are provided, so as to meet customer needs, reduce shopping time and increase customers' satisfaction and trust for companies. The above meticulous design of web portal can improve customer satisfaction and corporate competitiveness.

References

1. Bell, G., Gemmill, J.: Information superhighway dream. *Communications of the ACM* 39(7), 55 (1996)
2. Liu, C., et al.: Web sites of the Fortune 500 companies: facing customers through home pages. *Information & Management* 31(6), 335–345 (1997)
3. Rao Kowtha, N., Whai Ip Choon, T.: Determinants of website development: a study of electronic commerce in Singapore. *Information & Management* 39(3), 227–242 (2001)
4. Huizingh, E.K.: The content and design of web sites: an empirical study. *Information & Management* 37(3), 123–134 (2000)
5. Ghose, S., Dou, W.: Interactive functions and their impacts on the appeal of Internet presence sites. *Journal of Advertising Research* 38, 29–44 (1998)
6. Murtaza, A.H.: A framework for developing enterprise data warehouses. *Information Systems Management* 15(4), 21–26 (1998)
7. Angehrn, A.: Designing mature Internet business strategies: the ICDT model. *European Management Journal* 15(4), 361–369 (1997)
8. DeLone, W.H., McLean, E.R.: Information systems success: the quest for the dependent variable. *Information Systems Research* 3(1), 60–95 (1992)
9. Deighton, J.: The future of interactive marketing. *Harvard Business Review* 74(6), 151–161 (1996)
10. Seybold, P.B., Marshak, R.T.: *Customers. com: how to create a profitable business strategy for the Internet and beyond 1998: Times Business* (1998)
11. Saaty, T.: Decision making with dependence and feedback: The analytic network process, vol. 17. RWS Publications, Pittsburgh (1996)
12. Meade, L.M., Presley, A.: R&D project selection using the analytic network process. *IEEE Transactions on Engineering Management* 49(1), 59–66 (2002)
13. Saaty, T.L., Vargas, L.G.: Decision making with the analytic network process. Springer (2006)
14. Saaty, T.L., Takizawa, M.: Dependence and independence: From linear hierarchies to nonlinear networks. *European Journal of Operational Research* 26(2), 229–237 (1986)
15. Lee, J.W., Kim, S.H.: Using analytic network process and goal programming for interdependent information system project selection. *Computers & Operations Research* 27(4), 367–382 (2000)
16. Karsak, E.E., Sozer, S., Alptekin, S.E.: Product planning in quality function deployment using a combined analytic network process and goal programming approach. *Computers & Industrial Engineering* 44(1), 171–190 (2003)

17. Chung, S.-H., Lee, A.H., Pearn, W.: Product mix optimization for semiconductor manufacturing based on AHP and ANP analysis. *The International Journal of Advanced Manufacturing Technology* 25(11-12), 1144–1156 (2005)
18. Bozdog, C.E., Kahraman, C., Ruan, D.: Fuzzy group decision making for selection among computer integrated manufacturing systems. *Computers in Industry* 51(1), 13–29 (2003)
19. Emblemstvag, J., Tønning, L.: Decision support in selecting maintenance organization. *Journal of Quality in Maintenance Engineering* 9(1), 11–24 (2003)
20. Tran, L.T., et al.: Integrated environmental assessment of the mid-Atlantic region with analytical network process. *Environmental Monitoring and Assessment* 94(1-3), 263–277 (2004)
21. Chung, S.-H., Lee, A.H., Pearn, W.-L.: Analytic network process (ANP) approach for product mix planning in semiconductor fabricator. *International Journal of Production Economics* 96(1), 15–36 (2005)
22. Liou, T.-S., Wang, M.-J.J.: Ranking fuzzy numbers with integral value. *Fuzzy Sets and Systems* 50(3), 247–255 (1992)

Mobile Learning Achievement from the Perspective of Self-efficacy: A Case Study of Basic Computer Concepts Course

Yuh-Ming Cheng*, Sheng-Huang Kuo, and E-Liang Cheng

Computer Science and Information Engineering, Shu Te University, Taiwan
cymer@stu.edu.tw

Abstract. The major purpose of the research is to analyze the effect of mobile learning course in BCC on self-efficacy of student regarding learning achievement. The research utilizes cross-sectional research design and uses structural questions as research tool for data gathering. Statistical methods included descriptive statistics, t-test, Pearson Correlation coefficient. The results obtained two conclusions, from such a results, this paper suggests a further research on the self-efficacy, mobile learning, and learning achievement.

Keywords: learning achievement, mobile learning, self-efficacy.

1 Introduction

Education is an important investment in personal growth and development of the country, in today's era of globalization the world situation pulsating fast change, fierce international competition, and thus be highly competitive education to cultivate talent, effectively enhance national strength, but the only way, therefore the world countries have to inject a lot of educational resources and funding, an attempt to improve the environment through support, encourage students actively involved in learning, increase learning effectiveness, to enhance the quality of student learning. Human survival in complex environments, often encounter different problems and challenges. However, how people adjust themselves to deal with the complex environmental changes are often concerned about the problem of psychological research. In educational psychology, many studies about human learning behavior, one of which is the development of performance from self-efficacy from social cognitive theory, and discuss its impact on human learning. Self-efficacy theory has a very wide range of application, has been deep into the human psyche, and many other areas of life, for human self-understanding and improve behavioral performance, and create a happy life made a great contribution. The main purpose of this study was to analyze the effect of mobile learning management course in basic computer concepts (BCC) on self-efficacy of student regarding learning achievement.

* Corresponding author.

2 Literature Study

2.1 Self-efficacy

Self-efficacy theory developed in the 1970s after it was used to explore contemporary psychology and its ability to explain human perception. According to Bandura [1] definition of self-efficacy, self-efficacy is the ability of people to organize themselves and perform the actions necessary to achieve a specific program performance judgment. Self-efficacy is a concept associated with the ability to refer the individual to cope or deal with environmental incidents effectiveness or efficiency, but also refers to itself as a form of thinking individual objects, is a complex psychological process, control of human motivation and behavior [2].

Conversely, low self-efficacy weak people, you often choose to escape or hide ostrich mentality to face reality and work hard to solve the problem, and its experience of failure, in turn, strengthen or weaken their negative self-efficacy, creating a vicious cycle [3].

Many studies have found that self-efficacy was positively correlated with learning outcomes[4] [5] [6]. Hence, high self-efficacy learner self-efficacy compared to the lower, generally have higher learning motivation, easy-to-attainable therefore higher learning and satisfaction [7].

2.2 Mobile Learning

Advancement of technology has driven a strong demand for more sophisticated teaching instruments, for example, computer application, video, and various different equipment [8]. Technological enhanced teaching and learning activities will improve students' cognitive for those abstraction and intricacies of subjects [9]. According to McEwan and Cairncross [10], multimedia has the potential to create a higher learning quality. Sharples [11] said, learning that the use of mobile telecommunications equipment that considers a student's ability to communicate. In addition, the learning that occurs informally using mobile learning also considered as learning.

Under the fast developments in mobile and computer technologies, new methods in this area have also emerged besides traditional ones, and, one of the methods used web based distance education is mobile learning [12]. As a result of this, mobile learning methods and especially Web-based e-Learning have gained importance [13].

Kici said [14], mobile learning can be defend as easy and flexible learning due to the fact that it occurs regardless of time and place using portable mobile devices. Mobile learning is the use of mobile devices and wireless networks, such as mobile communications equipment, with mobile learning e-Learning system that allows learners regardless of time, place restrictions, in order to enjoy the convenience of learning, immediate resistance and suitability [15].

3 Research Design and Implementation

3.1 Research Design and Research Framework

The research utilizes cross-sectional research design and uses structural questions as research tool for data gathering. The research objects are based on the students from university of technology in Southern area. In this study, a "one-group posttest-only design," the study of 60 students for BCC course. This class students to perform 12 weeks of learning, to use a mobile phone or tablet or notebook computer to do action learning to verify our proposed model. According to research goal, literature study and several years of teaching experiences of the researcher, the learning achievement framework for using self-efficacy of university students with respect to mobile learning is as shown in Figure 1.

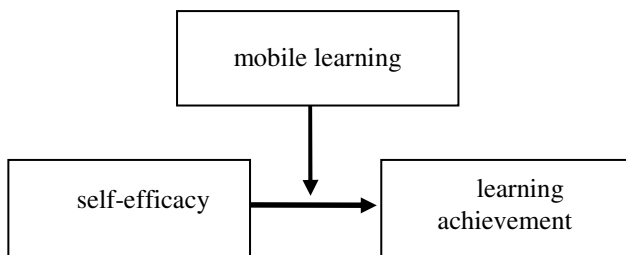


Fig. 1. Research Framework

3.2 Questionnaire Development of "Mobile Learning Achievement from the Perspective of Self-efficacy"

For the process of questionnaire development in the research, after literature is studied, initial questionnaire is drafted and modified based on questionnaire contents of using motivation for "Mobile Learning Achievement from the Perspective of Self-efficacy" proposed by 9 experts from interviews, followed by using Likert scale (five-point scale scoring) to give scores from 1 to 5 in terms of "strongly disagreed" to "strongly agreed".

There are 28 questionnaire items developed at first, followed by analyzing and arranging the scaling table. Subsequently, 7 domestic experts in self-efficacy field are invited to examine the contents of the scaling table repetitively. The words and sentences are modified and embellished in the questionnaire with respect to the applicability of each questionnaire item. There are total 17 questions in the scaling table after validity examination by experts.

3.3 Test of Confidence and Validity for Questionnaire

As for validity test of the research, the expert content validity is adopted for experts to check carefully whether the examination contents can represent the behavioral level to be measured. A pretest is performed for 30 students who have ever used mobile learning based on the purposive sampling approach, followed by performing the test of item

analysis to identify the contents in the scaling table. At last, the test of factor analysis is performed to acquire 3 aspects, which are named by "Self-efficacy", "Mobile Learning" and "Learning Achievement" according to the initially structured research variables. Then, the last question identification is performed to acquire 15 questions, including 5 "Self-efficacy" aspects, 5 "Mobile Learning" aspects and 5 "Learning Achievement" aspects, such that the framework validity of the scaling table is established.

After the scaling table is tested with respect to expert validity and framework validity, the test of confidence for the scaling table is performed directly with Cronbach α coefficient adopted to perform internal compliance analysis for questions under the same aspect. The Cronbach α coefficients for these aspects are 0.86 for "Self-efficacy", 0.88 for "Mobile Learning" and 0.90 for "Learning Achievement", respectively. The Cronbach α coefficient of the complete scaling table is 0.86. The confidences of the three aspects and the complete scaling table are all above at least 0.7, which is required by Nunnally [16].

4 Data Analysis and Result

After completed questionnaires are received, ineffective questionnaires are deleted to acquire really 60 questionnaires, which are performed with statistics and analysis by using SPSS 18.0 for Windows software. At first, the illustrative statistics is used to view the distribution for basic data of students. Next, independent sample t-test is used to test student genders. The major purpose is to see the scoring and whether significant difference is achieved with respect to "Mobile Learning Achievement from the Perspective of Self-efficacy". Then, Pearson Correlation Analysis is used to understand the correlation degree between learning achievement factors to use mobile learning from the perspective of self-efficacy.

4.1 Illustrative Statistics of Background Variables

From Table 1, there are total 60 test subjects, among which 35 are "males" occupying 58.3%, and 25 are "females" occupying 42.7%. From above data, in the "Gender" background variable, the numbers of test subjects in the two groups are very close.

Table 1. Table for Illustrative Statistics Summary of Different Background Variables

Background Variables	Group	Number of People	Proportion
Gender	Male	35	58.3%
	Female	25	42.7%

4.2 Independent Sample t-Test Analysis of Mobile Learning Achievement from the Perspective of Self-efficacy

From statistics result (as shown in table 2), the independent sample t-test of gender in the "Self-efficacy" "Self-efficacy" sub-scaling table does not achieve significant difference with $t(58)=-.305$, $p=.698$, 95%CI[-0.891,0.542]. Wherein the score with

respect to male (M=21.36) is not higher significantly than the core with respect to female (M=20.28) and the statistical power is 0.068. The independent sample t-test in the "Mobile Learning" sub-scaling table does not achieve significant difference with $t(58)=-.572$, $p=.498$, 95%CI[-1.211,0.653]. Wherein the score with respect to male (M=21.63) is not higher significantly than the core with respect to female (M=22.12) and the statistical power is 0.103. The independent sample t-test in the "Learning Achievement" sub-scaling table does not achieve significant difference with $t(58)=.751$, $p=.381$, 95%CI[-0.520,1.476]. Wherein the score with respect to male (M=22.10) is not higher significantly than the core with respect to female (M=20.73) and the statistical power is .128.

Table 2. Average t-Test of Individual for Mobile Learning Achievement from the Perspective of Self-efficacy

Variables	Male		Female		T Value	P Value	95% CI		η^2	1- β
	(n=35)		(n=25)				LL	UL		
	M	SD	M	SD						
Self-efficacy	21.36	2.62	20.28	2.35	-.305	.698	-0.891	0.542	.002	.068
Mobile Learning	21.63	2.84	22.12	3.52	-.572	.498	-1.211	0.653	.004	.103
Learning Achievement	22.10	3.32	20.73	3.82	.751	.381	-0.520	1.476	.005	.128

4.3 Pearson Correlation Analysis of Mobile Learning Achievement from the Perspective of Self-efficacy

In the entire research, Pearson Correlation Statistical Analysis is used at last to analyze product moment correlations between Self-efficacy, Mobile Learning and Learning Achievement of using Mobile Learning Achievement from the Perspective of Self-efficacy. There is significant low positive correlation between both. As for "Self-efficacy", "Mobile Learning" and "Learning Achievement" all have significantly high positive correlations with the overall aspect, showed as Table 3.

Table 3. Pearson Correlation Analysis of Mobile Learning Achievement from the Perspective of Self-efficacy

Aspects	Self-efficacy	Mobile Learning	Learning Achievement
Self-efficacy	-	-	-
Mobile Learning	.386***	-	-
Learning Achievement	.492***	.416***	-
Average	20.82	21.88	21.42
Standard Deviation	2.49	3.18	3.57

N=60, *** $p<.001$

5 Conclusion

The major purpose of the research is to analyze the effect of mobile learning course in BCC on self-efficacy of student regarding learning achievement. The major purpose of the research is to analyze the effect of mobile learning course in BBC on self-efficacy of student regarding learning achievement of university students with respect to the Mobile Learning Achievement from the Perspective of Self-efficacy, and make conclusion and propose specific recommendation according to the research results and discoveries.

The First, for the process of composing questionnaires of the research, after studying literatures, initial questionnaire is drafted, and is modified with respect to the questionnaire contents of Mobile Learning Achievement from the Perspective of Self-efficacy from interviews with 9 experts. Likert scale (five-point scale scoring) is adopted to design "Mobile Learning Achievement from the Perspective of Self-efficacy" scaling table. There are 15 questions acquired from tests of expert validity and framework validity. The acquired aspects are "Self-efficacy", "Mobile Learning" and "Learning Achievement", respectively. Then, the internal compliance analysis is performed. Cronbach α coefficients of all aspects are all above standard. The overall scaling table confidence is good. This scaling table may be used to measure the Mobile Learning Achievement from the Perspective of Self-efficacy in the future.

The second, self-efficacy, Mobile Learning and Learning Achievement there are a positive correlation between the each other. The results of this study are similar to studies with Reisherei [17], Reisherei said, "Self- efficacy includes:" person's belief to his capabilities in organize and implement necessary activities for manage different conditions and situations". In other words, self- efficacy is person's belief to his ability to success in a specific situation [18]. According to Bandura's self- efficacy theory, behavior change and maintenance are function of expectations about the outcome and belief in one's ability to engage in or execute the behavior (efficacy) ". In particular, the study of the integration of self-efficacy in mobile learning, Dykes and Knight [19] proposed, Implementing mobile services in education in the form of mobile learning modules is an innovative process at many levels of higher education. Also, Oberer and Erkollar [20] pointed out, mobile learning can be used to enhance the overall learning experience of students and teachers, and, through mobile support learners' throughput rates might be improved and the quality of the learning experience enhanced. Hence, given the growing use of mobile devices, there is now increasing interest in the potential for supporting the mobile learner [21].

Acknowledgements. The authors greatly appreciate the financial support provided by the National Science Council, Taiwan, ROC, under contract No. NSC 102-2511-S-366-001-MY2.

References

1. Bandura, A.: Social foundations of thought and action: A social cognitive theory. Prentice-Hall, Inc., Englewood Cliffs (1986)
2. Cheng, T.-F., Wu, H.-C.: A study of the Relationships in Clerk's Self-efficacy, Organizational Commitment, and Job Satisfaction in Different Types of Paternalistic Leadership: An Application of Structural Equation Modeling. *Journal of Education and Psychology* 29(1), 47–75 (2006)
3. Lee, Z.-P.: A Study of Improving the Learning Outcomes of Ceramics Skills with Self-Efficacy Strategy. *Art Journal* 83, 37–57 (2008)
4. Karagüven, M.H., Yukselöglü, S.M.: Vocational Self-efficacy and Academic Motivation Levels of Technical and Vocational Pre-service Teachers (Example of Marmara University). *Procedia - Social and Behavioral Sciences* 106, 3366–3374 (2013)
5. Karnell, A.P., Cupp, P.K., Zimmerman, R.S., Feist-Price, S., Bennie, T.: Efficacy of an american alcohol and hiv prevention curriculum adapted for use in south africa: Results of a pilot study in five township schools. *AIDS Education and Prevention* 18(4), 295–310 (2006)
6. Lent, R.W., Lopez, F.G., Bieschke, K.J.: Predicting mathematics related choice and success behaviors: Test of an expanded social cognitive model. *Journal of Vocational Behavior* 42(2), 223–236 (1993)
7. Huang, J.-T., Wang, Y.-C., Li, T.-C.: A Study of the Relationship of Employee Self-Efficacy, Learning Strategy, and E-Learning Effectiveness. *K.U.A.S. Journal of Humanities and Social Sciences* 6(2), 119–142 (2009)
8. Irwan, I.M., Norazah, M.N., Ridzwan, C.R., Rosseni, D.: The Acceptance of AutoCAD Student for Polytechnic on Mobile Learning. *Procedia - Social and Behavioral Sciences* 102(22), 169–176 (2013)
9. Ahmad Rizal, M., Yahya, B.: The Effect of Courseware Utilization to the Student's Achievement for Field Independencedependence Cognitive Styles Student. *Journal of Technical, Vocational & Engineering Education* 4, 12–21 (2011)
10. McEwan, T., Cairncross, S.: Evaluation and multimedia learning objects: towards a human-centred approach. *Interactive Technology and Smart Education* 1(2), 101–112 (2004)
11. Sharples, M.: Big issues in mobile learning. Report of a workshop by the Kaleidoscope Network of Excellence Mobile Learning Initiative. University of Nottingham, UK (2006)
12. Göksu, İ., Atici, B.: Need for Mobile Learning: Technologies and Opportunities. *Procedia - Social and Behavioral Sciences* 103(26), 685–694 (2013)
13. Yildirim, S., Goktas, Y., Temur, N., Kocaman, A.: A Checklist for a Good Learning Management System (LMS). *Türk Egitim Bilimleri Dergisi* 4(2) (2004) (in Turkish), http://www.tebd.gazi.edu.tr/arsiv/2004_cilt2/sayi_4/455-462.pdf
14. Kici, D.: A Study on the Effects of expectations for Mobile Learning University Education of University Students. In: *International Conference on New Trends in Education and Their Implications (ICONTE)*, Antalya-Turkey, November 11-13 (2010)
15. Chang, C.-C., Lin, C.-L., Yan, C.-F.: The Influence of Perceived Convenience and Curiosity on Continuous English Learning Intention in Mobile Environment. *Journal of Educational Media & Library Sciences* 48(4), 571–588 (2011)
16. Nunnally, J.C.: *Psychometric theory*. McGraw-Hill, New York (1978)

17. Reishehrei, A.P., Reishehrei, A.P., Soleimani, E.: A Comparison Study of Self Concept and Self Efficacy in Martial Arts and non Martial Arts Athletics in Iran. *Procedia - Social and Behavioral Sciences* 116, 5025–5029 (2014)
18. Bandura, A.: Social cognitive theory of self- regulation. *Organizational Behavior and Human Decision Processes* 50, 248–287 (1991)
19. Dykes, G., Knight, H.: Mobile learning for teachers in Europe. Exploring the potential of mobile technologies to support teachers and improve practices. UNESCO Working Paper Series on Mobile Learning, France (2012)
20. Oberer, B., Erkollar, A.: Mobile Learning in Higher Education: A Marketing Course Design Project in Austria. *Procedia - Social and Behavioral Sciences* 93, 2125–2129 (2013)
21. Li, L., Leina, L.: Designing Principles of Mobile Learning in ESP Course for Chinese Students. *IERI Procedia* 2, 142–148 (2012)

Part VII
Intelligent Technologies and Telematics
Applications

The Implementation of OBD-II Vehicle Diagnosis System Integrated with Cloud Computation Technology

Jheng-Syu Jhou and Shi-Huang Chen

Department of Computer Science & Information Engineering, Shu-Te University
No.59, Hengshan Rd., Yanchao Dist., Kaohsiung City
82445, Taiwan (R.O.C.)
shchen@stu.edu.tw

Abstract. This paper implemented a cloud computation based second generation on-board diagnostic (OBD-II) system. The proposed system is integrated with OBD-II, 3.5G wireless network, and cloud computing technologies. It can perform real-time vehicle status surveillance. The monitored features cover engine rpm, vehicle speed, coolant temperature, fault codes, and other vehicle dynamics information. The vehicle information will be transmitted to the cloud computing server via 3.5G wireless network for fault analysis. Once cloud computing server detects fault conditions, the proposed system could classify the fault conditions depended on vehicle type and its model year. Then the cloud computing server will report the fault code analysis results to the user and provide the description about repair procedure. The proposed system will greatly shorten the time to detect vehicle trouble condition. The system presented in this thesis has a very high value in the applications of vehicle maintenance and fleet management.

Keywords: On board diagnostics (OBD), cloud computing, vehicle repair information, android system.

1 Introduction

The major vehicle companies have taken the durability of these automotive electronic devices into account when designing modern vehicles, however, the human failure or improper operation will still lead to unnecessary fuel consumption and exhaust pollution. Because of these modern vehicles equipped with lots of electronic components, it is not easy to diagnose these vehicle faults using traditional fault detection methods. According to previous researches [1], the time for finding vehicle fault is 70%, while the time for troubleshooting and maintenance accounts is just 30%. Therefore, the major vehicle companies developed a fault diagnosis system, namely, on-board diagnostic or OBD, into vehicle electronic control unit (ECU).

The major vehicle companies have taken the durability of these automotive electronic devices into account when designing modern vehicles, however, the human failure or improper operation will still lead to unnecessary fuel consumption and exhaust pollution. Because of these modern vehicles equipped with lots of electronic components, it is not easy to diagnose these vehicle faults using traditional fault

detection methods. According to previous researches [1], the time for finding vehicle fault is 70%, while the time for troubleshooting and maintenance accounts is just 30%. Therefore, the major vehicle companies developed a fault diagnosis system, namely, on-board diagnostic or OBD, into vehicle electronic control unit (ECU).



Fig. 1. Various symbols of MIL or the check engine light. (Pictures are copies from Ford, GM, and Toyota vehicle maintain manuals).

The OBD system is designed to consecutively monitor the running condition of vehicle [2]-[4]. Once there is a malfunctioning element that controls the emission of exhaust, the OBD system will turn on the Malfunction Indicator Lamp (MIL) or the Check Engine light, as shown in Figure 1, to notify the driver to repair the vehicle immediately. When the OBD system detects malfunctions, OBD regulations will inform the ECU of the vehicle to save a standardized Diagnostic Trouble Code (DTC) about the information of malfunctions in the memory. An OBD Scan Tool for the servicemen can access the DTC from the ECU to quickly and accurately confirm the malfunctioning characteristics and location in accordance with the prompts of DTC. In addition to DTC, the OBD system can monitor more than 80 items of real-time driving status, e.g., vehicle speed, engine rpm, throttle position, intake air temperature, engine coolant temperature, and etc [2]-[4].

The OBD system is widely used in the current vehicle workshops or service dealers. Due to the operation of OBD is quit difficult, the general drivers could not access OBD data easily. Hence, this paper develops a vehicle diagnosis system integrated with cloud computation technology to help driver or repairer to determine the faults of vehicle. The vehicle diagnosis system proposed in this paper consists of on-board unit (OBU) and vehicle diagnostic server (VDS). The OBU is divided into OBD and CAN Bus signal receptions, GPS receiver, and 3.5G wireless network module. In addition to diagnostic server monitoring function, the proposed VDS integrates the online expert system and statistical analysis to strengthen the functionality.

The remaining sections of this paper are organized as follows. Sections II and III will introduce the OBU and VDS, respectively. Section IV describes the experimental results. The final section will conclude this paper.

2 On-Board Unit (OBU)

The OBU module proposed in this paper is designed to acquire the real-time vehicle location and operation information, e.g., date, time, longitude, latitude, speed, engine rpm, coolant temperature, fault code number, and others from GPS receiver and CAN/OBD-II adapter. The real-time vehicle information will be encoded via CAN/OBD-II diagnosis encoder and then transferred to digital bit streams. The OBU module will transmit these digital bit-streams to the vehicle diagnostic server (VDS) through the 3.5G network. Figure 2 is the block diagram of the proposed OBU module. The system is mainly comprised of CAN/OBD-II adapter, GPS receiver,

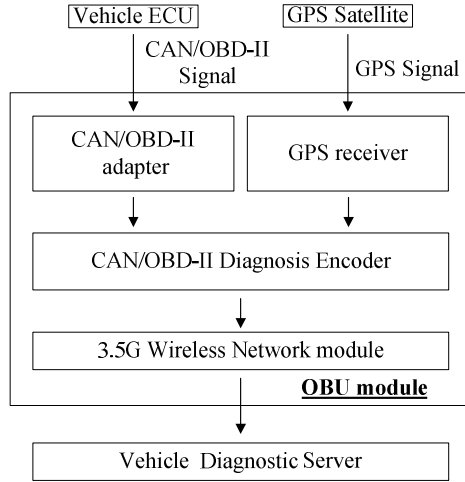


Fig. 2. The block diagram of the proposed OBU module

CAN/OBD-II diagnosis encoder, and 3.5G wireless network module. The following three subsections will particularly introduce these three main items.

2.1 CAN/OBD-II Adapter

The CAN/OBD-II adapter used in this paper is based on ELM 327 chip and follows SAE/ISO standards [6]-[9], [12]. The main features of CAN/OBD-II are (1) unified J1962 16-pin socket and data link connector (DLC) (as shown in Figure 3); (2) unified DTC and meanings; (3) storage and display DTC; (4) vehicle record capability; and (5) auto-clear or reset function for the DTC. In other words, just one set of CAN/OBD-II scan tool is able to perform the diagnosis task and can scan against variety of vehicles which equipped with CAN/OBD-II system.



Fig. 3. (a) J1962 CAN/OBD-II 16-pin socket, (b) CAN/OBD-II DLC

There are five codes in total to represent the OBD-II DTC message. Figure 4 shows the definition of the OBD-II DTC. The first code is an English alphabet to stand for the established malfunction system. The remaining four codes are digits; the second code indicates the meaning of malfunction formulated by ISO/SAE or customized by the vehicle manufacturer; the third code shows the area of vehicle system; the remaining two codes represent the definition of the subject malfunction [5].

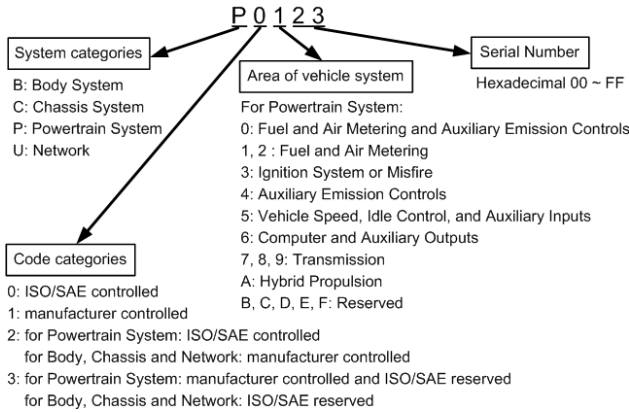


Fig. 4. Definition of the CAN/OBD-II diagnostic trouble code (DTC)

2.2 GPS Receiver

GPS receiver can receive signals from 8-12 sets of GPS satellite at the same time. The GPS satellite signals include coordinated universal time, ephemeris data, almanac data, coarse/acquisition code, and etc. GPS receiver can receive, process, and transform the information into time, latitude, longitude, velocity, orientation, altitude, estimated position error, and other time-location information [9]. Then, these data will be transmitted to a Geographical Information System (GIS), such as Google Maps to pinpoint and display the vehicle position. The proposed system used the GPS receiver with the SiRF Star III chipset mounted to collect the GPS signals.

2.3 CAN/OBD-II Diagnosis Encoder

The function of CAN/OBD-II diagnosis encoder is to encode and integrate the GPS signals as well as CAN/OBD information in accordance with the preset transmission format. These encoded digital bit streams will be transmitted to the vehicle diagnostic server via 3.5G wireless network [10]. The vehicle diagnostic server can decode the digital bit streams in accordance with the predefined transmission format to acquire the subject vehicle information, including speed, engine rpm, engine coolant temperature, OBD DTC, and the GPS coordinates for the position of vehicle.

3 Vehicle Diagnostic Server (VDS)

The VDS proposed in this paper could be regarded as a kind of vehicle diagnostics management platform. It will receive the real-time vehicle data from OBU over the wireless network. These real-time vehicle data include GPS coordinates, vehicle speed, engine rpm, coolant temperature, OBD DTC, and etc. The VDS will analyze these real-time vehicle data using online expert system with statistical analysis. Once the online expert system detects any vehicle abnormal condition, the VDS will immediately notify the driver for the necessary repair.

The built-in online expert system will analyze the vehicle speed, engine rpm, throttle angle, brake, engine temperature, battery voltage, oxygen sensor voltage, fuel injection frequency, instantaneous fuel consumption, and other information by statistical algorithms to determine whether the vehicle is abnormal and perform vehicle fault warning. The VDS also supports the following data management functions: the driver and vehicle data management, real-time vehicle location communications, vehicle running state, the vehicle exception alerts, remote vehicle maintenance instructions, and etc.

3.1 Online Expert System

Generally, the expert system is designed for a particular field to solve, judge, or explain a problem through a knowledge database and an inference engine. With the gradual development of a variety of expert systems, the expert systems have been used in the various industries, and more diverse applications.

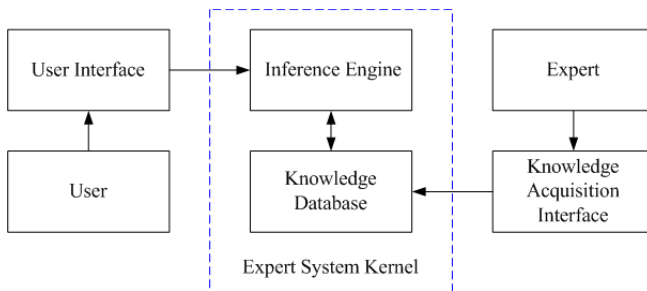


Fig. 5. The block diagram of the expert system

In this paper, the online expert system is mainly designed to enhance the functionality of the VDS. The online expert system contains a knowledge database, knowledge acquisition interface, user interface, and inference engine. Figure 5 shows the block diagram of the expert system.

The knowledge database is built to collect human expert knowledge in systematically express or modular. Hence the computer can carry out inferences and solve the problem. The purpose of knowledge acquisition interface is to help system developers to conduct knowledge extraction, editing and revision of knowledge database and inferences device. It also can test, record, and describe the status as well as results of the expert system. The user interface is a bridge of communication between user and expert system. The design of user interface focuses on compatibility and convenience. It usually provides a variety of methods of operation and indicates the correct behavior patterns.

The inferences engine is the main part of the expert system. It makes use of algorithms or decision-making strategies to conduct the expertise and knowledge database of inferences between knowledge. Common expert system could be divided into rule-based reasoning and case-based management system.

The rule-based inferences engine uses plenty of rules built into the knowledge database. The reasoning process is mainly by the rules of the link, the system must control the initial state to reach its goal of reasoning. Therefore, rule-based expert system is suitable for a known complete range or narrower application. Generally, the rule-based expert system includes three steps:

1. Matching: to find all the rules in line that meet the case.
2. Selection: In all candidate rules, decided one of the most suitable choice of the status of execution.
3. Rule execution: To execute the selection rules of the process.

The case-based inferences engine is derived from machine learning domain. It is essentially a human problem-solving mode. The basic approach is to use the past similar experience to solve new problems. As long as the collection of adequate fault case is enough, the reliability of system will be increase. Case-based expert system has following five main steps.

1. Store the related examples or experiences in a knowledge database.
2. Identify and understand the current problems.
3. Extract the similar case of example or experience from the knowledge database.
4. Use these similar examples or experiences to solve current problems.
5. Once the problem is solved, update this example or experience in the knowledge database.

Since the entire vehicle system composed of various sub-systems, the event of failure may quite different. The cause of the failure may simply occur in the same subsystem. The failure of explicit phenomenon can be judged as a sub-system failure, and it is suitable for rule-based inferences engine. However, the fault may fail to cover the number of sub-systems. It usually does not have explicit symptom and is suitable for use case-type inferences engine. Therefore, this paper will use rule-based and case-based simultaneously to design the expert system.

4 Experimental Results

This paper used two notebook computers to simulate the OBU and VDS modules. OBU module has a GPS-1155 (USB interface, SiRF III) GPS receiver and a Bluetooth receiver to get the GSP signal and CAN/OBD-II information, respectively. It also comes with a 3.5G network module. The notebook used to simulate VDS installs Microsoft Access 2003 to record the real-time vehicle information and makes use of Google Maps API to indicate the vehicle location in the GIS. VDS has a fix IP address to receive the vehicle information send by OBU module. All of the subsystems proposed in this paper are implemented by C# language and the computer is able to execute this browser when it is installed the Microsoft .NET Framework V2.0. Figure 6 shows the simulation equipments mentioned above.



Fig. 6. The simulation equipments of the proposed system

Figure 7 shows the CAN/OBD-II information displayer implemented in this paper. In this example the temperature of engine coolant, i.e. 164 degree, exceeds the legal limit and hence the VDS will notify the driver as shown in Figure 7.



Fig. 7. The CAN/OBD-II information displayer implemented in this paper. They are vehicle speed, engine rpm, temperature of engine coolant, MAF (Mass Air Flow), ECU voltage.

Once VDS detects any fault of vehicle, the VDS will first determine the cause of this fault using the on-line expert system. Then OBU can show the service manual downloaded from VDS. Figure 8 is an example of on-line service manual provided by VDS.



Fig. 8. The on-line repair manual provided by VDS

5 Conclusion

This paper integrated OBD-II system, 3.5G mobile network, GPS, and cloud computing technologies to implement an OBD-II vehicle diagnosis system for real-time vehicle status surveillance applications. The proposed system is comprised of on-board unit (OBU) module and vehicle diagnostics server (VDS). The OBU will monitor engine rpm, vehicle speed, coolant temperature, fault codes, and other vehicle dynamics information. The vehicle information will be transmitted to the cloud computing server, i.e., VDS, via 3.5G wireless network for fault analysis. Once VDS detects fault conditions, the proposed system could classify the fault conditions depended on vehicle type and its model year. Then VDS will report the fault code analysis results to the user and provide the description about repair procedure. The proposed system will greatly shorten the time to detect vehicle trouble condition. The system presented in this thesis has a very high value in the applications of vehicle maintenance and fleet management.

References

1. Jie, H., Fuwu, Y., Jing, T., Pan, W., Kai, C.: Developing PC-Based Automobile Diagnostic System Based on OBD System. In: 2010 Asia-Pacific Power and Energy Engineering Conference (APPEEC), pp. 1–5 (March 2010)
2. Diagnostic Trouble Code Definitions Equivalent to ISO/DIS 15031-6, SAE Standard J2012 (2002)
3. E/E Diagnostic Test Modes — Equivalent to ISO/DIS 15031-5, SAE Standard J1979 (2002)
4. Diagnostic Connector Equivalent to ISO/DIS 15031-3, SAE Standard J1962 (2002)
5. Diagnostic Trouble Code Definitions Equivalent to ISO/DIS 15031-6, SAE Standard J2012 (2002)
6. E/E Diagnostic Test Modes — Equivalent to ISO/DIS 15031-5, SAE Standard J1979 (2002)
7. Lin, C.E., Shiao, Y.-S., Li, C.-C., Yang, S.-H., Lin, S.-H., Lin, C.-Y.: Real-Time Remote Onboard Diagnostics Using Embedded GPRS Surveillance Technology. *IEEE Trans. on Vehicular Technology* 56(3), 1108–1118 (2007)
8. Lin, C.E., Li, C.C., Yang, S.H., Lin, S.H., Lin, C.Y.: Development of On-Line Diagnostics and Real Time Early Warning System for Vehicles. In: *Proc. IEEE Sensors for Industry Conference*, pp. 45–51 (February 2005)
9. NMEA data, <http://gpsinformation.org/daled/nmea.htm>
10. Holma, H., Toskala, A.: *HSDPA/HSUPA for UMTS - High Speed Radio Access for Mobile Communications*. Wiley, John & Sons Ltd. (2006)

Daily Power Demand Forecast Models of the Differential Polynomial Neural Network

Ladislav Zjavka

VŠB-Technical University of Ostrava, IT4innovations Ostrava, Czech Republic
lzjavka@gmail.com

Abstract. Short-term electric energy estimations of a future demand are needful for the planning of generating electricity of regional grid systems and operating power systems. In order to guarantee a regular supply, it is necessary to keep a reserve. However an over-estimating of a future load results in an unused spinning reserve. Under-estimating a future load is equally detrimental because buying at the last minute from other suppliers is obviously too expensive. Cooperation on the electricity grid requires from all providers to foresee the demands within a sufficient accuracy. Differential polynomial neural network is a new neural network type, which forms and solves an unknown general partial differential equation of an approximation of a searched function, described by data observations. It generates convergent sum series of fractional polynomial derivative terms. This operating principle differs by far from other common neural network techniques. In the case of a prediction of only 1-parametric function, described by real data time-series, an ordinary differential equation is constructed and substituted with partial derivatives.

Keywords: power demand prediction, week load cycle, differential polynomial neural network, sum relative derivative term.

1 Introduction

Electricity demand accurate forecasts are necessary in the operation of electric power system, producers should develop strategies to maximize the profits and minimize risks. Day-ahead load forecasting techniques can involve autoregressive integrated moving average (ARIMA) [3], chaotic dynamic non-linear models with evolutionary hybrid computation [9], exponential smoothing or interval time-series [4]. A real data 1-parametric function time-series progress is difficult to predict using deterministic methods as weather conditions as well as other extraneous factors can by far influence it. The power demand model could be trained with data relations of several subsequent days in previous weeks at the same time points, as the daily power cycles of each following week are of a similar progress. After that the prediction is formed with respect to few last days with the same denomination in the current week. The model should be updated to take into account a dynamic character of the problem. Artificial neural network (ANN) is a powerful tool to deal with problems, which other method

solutions fail. It can define simple and reliable models, which exact solution is problematic or impossible to get using standard regression techniques.

$$y = a_0 + \sum_{i=1}^m a_i x_i + \sum_{i=1}^m \sum_{j=1}^m a_{ij} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m a_{ijk} x_i x_j x_k + \dots \tag{1}$$

m – number of variables $X(x_1, x_2, \dots, x_m)$ $A(a_1, a_2, \dots, a_m), \dots$ - vectors of parameters

Differential polynomial neural network (D-PNN) is a new neural network type designed by the author, which results from the GMDH (Group Method of Data Handling) polynomial neural network (PNN), created by a Ukrainian scientist Aleksey Ivakhnenko in 1968, when the back-propagation technique was not known yet. It is possible to express a general connection between input and output variables by the Volterra functional series, a discrete analogue of which is Kolmogorov-Gabor polynomial (1). This polynomial can approximate any stationary random sequence of observations and can be computed by either adaptive methods or system of Gaussian normal equations. GMDH decomposes the complexity of a process into many simpler relationships each described by the low order polynomials (2) for every pair of the input values. Typical GMDH network maps a vector input \mathbf{x} to a scalar output y , which is an estimate of the true function $f(\mathbf{x}) = y'$ [7].

$$y = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \tag{2}$$

D-PNN can combine the PNN functionality with some math techniques of differential equation (DE) solutions. Its models lie on the boundary of neural networks and exact computational techniques. D-PNN forms and resolves an unknown general DE description of an approximation of a searched function. It produces sum series of fractional polynomial derivative terms, which substitute for a DE, decomposing a system model into many partial derivative specifications of data relations. In contrast with the ANN functionality, each neuron (i.e. derivative term) can take part directly in the total network output calculation, which is generated by the sum of the active neuron output values [11].

2 Forecasting Energy Demands

The potential benefits of an energy demand prediction are obvious useful in automatic power dispatch, load scheduling and energy control. Load forecasting is important for economically efficient operation and effective control of power systems and enables to plan the load of generating unit. The purpose of the short-term electricity demand forecasting is to forecast in advance the system load, represented by the sum of all consumers load at the same time. A precise load forecasting is required to avoid high generation cost and the spinning reserve capacity. Under-prediction of the demands leads to an insufficient reserve capacity preparation and can threaten the system stability, on the other hand, over-prediction leads to an unnecessarily large reserve that leads to a high cost preparations. The nature of parameters that affect this problem includes many uncertainties. The accuracy of a dispatching system is influenced by

various conditional input parameters (weather, time, historical data and random disturbances), which can a prediction model involve, applying e.g. fuzzy logic [6]. ANN is able to model the non-linear nature of dynamic processes, reproduce an empirical relationship between some inputs and one or more outputs. It is applied for such purpose regarding to its approximation capability of any continuous nonlinear function with arbitrary accuracy that offer an effective alternative to more traditional statistical techniques. The load at a given hour is dependent not only on the load at the previous hour, but also on the load at the same hour on the previous day, and on the load at the same hour on the day with the same denomination in the previous week. There are also many important exogenous variables that should be considered, especially weather-related variables [5].

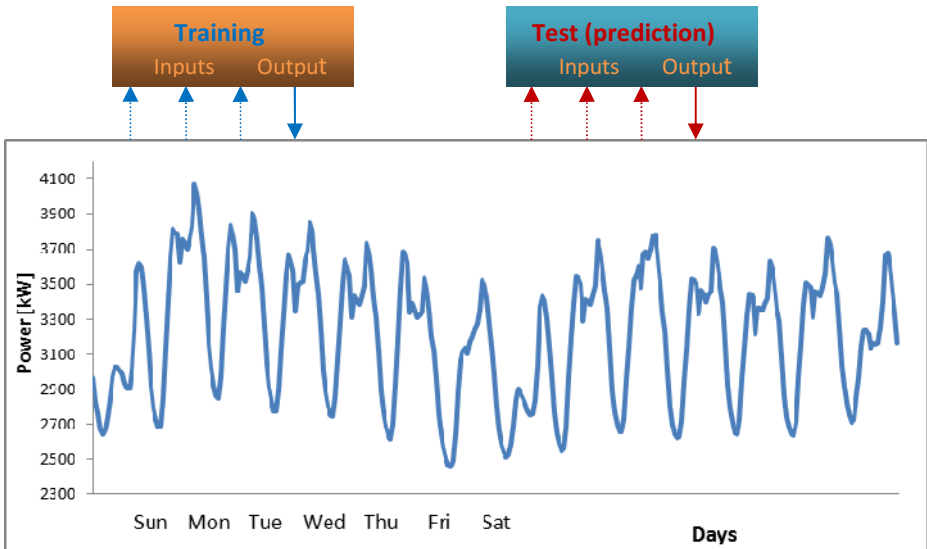


Fig. 1. Typical week cycles of power demands (March-April) [11]

The proposed method keynote of the power demand forecasting (using only 1-parametric function time-series) is to train a neural network model with daily cycle similarity relations of previous weeks, concerning several consecutive days with the same denomination, foregoing the 24-hour prediction. Power values of the 3 consequent days in previous weeks at the same time points form the input vector while the following day 24-hour shifted series define desired network outputs of the training data set. After training the network can estimate the following day power progress, using 3 input vector variables of 24-hour shifted time-series of the same time stamps of the current week last 3 days (Fig.1.). The model does not allow for weather or other disturbing effects, as these are not at disposal in the majority of cases. The power demand day cycle progress is much more variable in winter than summer months, which is influenced largely by using heating systems and temperature conditions [11].

3 General Differential Equation Composition

The D-PNN decomposes and substitutes for a general sum partial differential equation (3), in which an exact definition is not known in advance and which can generally describe a system model, with a sum of relative multi-parametric polynomial derivative convergent term series (4).

$$a + \sum_{i=1}^n b_i \frac{\partial u}{\partial x_i} + \sum_{i=1}^n \sum_{j=1}^n c_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} + \dots = 0 \quad u = \sum_{k=1}^{\infty} u_k \quad (3)$$

u = f(x₁, x₂, ..., x_n) – searched function of all input variables
a, B(b₁, b₂, ..., b_n), C(c₁₁, c₁₂, ...) – polynomial parameters

Partial DE terms are formed according to the adapted method of integral analogues, which is a part of the similarity model analysis. It replaces mathematical operators and symbols of a DE by the ratio of the corresponding values. Derivatives are replaced by their integral analogues, i.e. derivative operators are removed and simultaneously along with all operators are replaced by similarly or proportion signs in equations to form dimensionless groups of variables [2].

$$u_i = \frac{(a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2 + a_4 x_1^2 + a_5 x_2^2 + \dots)^{m/n}}{b_0 + b_1 x_1 + \dots} = \frac{\partial^m f(x_1, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_m} \quad (4)$$

n – combination degree of a complete polynomial of n-variables
m – combination degree of denominator variables

The fractional polynomials (4), which substitute for the DE terms, describe partial relative derivative dependent changes of *n*-input variables. The numerator (4) is a polynomial of *n*-input variables and can partly define an unknown function *u* of eq. (3). The denominator includes an incomplete polynomial of the competent derivative combination of variables. The root function of the numerator decreases a combination degree of the input polynomial of a term (4), in order to get the dimensionless values [2]. In the case of time-series data observations an ordinary differential equation (6) is formed with time derivatives describing 1-parametric function progress using partial DE terms, analogous to the general partial DE (3) construction.

$$a + bf + \sum_{i=1}^m c_i \frac{df(t, x_i)}{dt} + \sum_{i=1}^m \sum_{j=1}^m d_{ij} \frac{d^2 f(t, x_i, x_j)}{dt^2} + \dots = 0 \quad (5)$$

f(t, x) – function of time t and independent input variables x(x₁, x₂, ..., x_m)

Blocks of the D-PNN (Fig.2.) consist of derivative neurons having the same inputs, one for each fractional polynomial derivative combination, so each neuron is considered a summation DE term (4). Each block contains a single output GMDH polynomial (2), without derivative part. Neurons do not affect the block output but can participate directly in the total network output sum calculation of a DE composition. Each block has *l* and neuron 2 vectors of adjustable parameters *a*, then *a, b*.

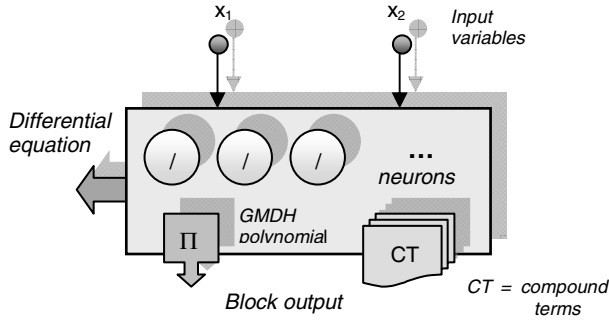


Fig. 2. D-PNN block involves basic and compound neurons (DE terms)

When in use 2 input variables the 2nd order partial DE is formed (6), which involves derivative of all the variables of the GMDH polynomial (2). D-PNN blocks form 5 simple neurons, which substitute for the DE terms in respect of derivatives of the single x_1, x_2 square x_1^2, x_2^2 and combination x_1x_2 variables of the 2nd order partial DE (6) solution, most often used to model physical or natural system non-linearities.

$$F\left(x_1, x_2, u, \frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, \frac{\partial^2 u}{\partial x_1^2}, \frac{\partial^2 u}{\partial x_1 \partial x_2}, \frac{\partial^2 u}{\partial x_2^2}\right) = 0 \tag{6}$$

where $F(x_1, x_2, u, p, q, r, s, t)$ is a function of 8 variables

4 Differential Polynomial Neural Network

Multi-layer networks forms composite polynomial functions (Fig.3.). Compound terms (CT), i.e. derivatives in respect of variables of previous layer blocks, are calculated according to the composite function partial derivation rules (7)(8). They are formed by products of the partial derivatives of external and internal functions.

$$F(x_1, x_2, \dots, x_n) = f(y_1, y_2, \dots, y_m) = f(\phi_1(X), \phi_2(X), \dots, \phi_m(X)) \tag{7}$$

$$\frac{\partial F}{\partial x_k} = \sum_{i=1}^m \frac{\partial f(y_1, y_2, \dots, y_m)}{\partial y_i} \cdot \frac{\partial \phi_i(X)}{\partial x_k} \quad k=1, \dots, n \tag{8}$$

All blocks contain 5 simple neurons, e.g. the 1st block of the last (3rd) hidden layer forms linear CT (9). The blocks of the 2nd and following hidden layers are additionally extended with CT (neurons), which form composite derivatives with respect to the output and input variables of the back connected previous layer blocks, e.g. (10)(11). The number of neurons of blocks, which involve composite function derivatives, doubles each previous back-connected layer [11].

$$y_1 = \frac{\partial f(x_{21}, x_{22})}{\partial x_{21}} = w_1 \frac{(a_0 + a_1x_{21} + a_2x_{22} + a_3x_{21}x_{22} + a_4x_{21}^2 + a_5x_{22}^2)^{\sqrt{2}}}{1.5 \cdot (b_0 + b_1x_{21})} \tag{9}$$

$$y_2 = \frac{\partial f(x_{21}, x_{22})}{\partial x_{11}} = w_2 \frac{(a_0 + a_1 x_{21} + a_2 x_{22} + a_3 x_{21} x_{22} + a_4 x_{21}^2 + a_5 x_{22}^2)^{\frac{1}{2}}}{1.6 \cdot x_{22}} \cdot \frac{(x_{21})^{\frac{1}{2}}}{1.5 \cdot (b_0 + b_1 x_{11})} \quad (10)$$

$$y_3 = \frac{\partial f(x_{21}, x_{22})}{\partial x_{11}} = w_3 \frac{(a_0 + a_1 x_{21} + a_2 x_{22} + a_3 x_{21} x_{22} + a_4 x_{21}^2 + a_5 x_{22}^2)^{\frac{1}{2}}}{1.6 \cdot x_{22}} \cdot \frac{(x_{21})^{\frac{1}{2}}}{1.6 \cdot x_{12}} \cdot \frac{(x_{11})^{\frac{1}{2}}}{1.5 \cdot (b_0 + b_1 x_{11})} \quad (11)$$

The square (12) and combination (13) derivative terms are also calculated according to the composite function derivation rules. The denominator coefficients balance a length variety of the derivative polynomials (11)(12)(13). D-PNN is trained with only a small set of input-output data samples, in a similar way to the GMDH algorithm [7]. The number of network hidden layers coincides with a total number of input variables.

$$y_4 = \frac{\partial^2 f(x_{21}, x_{22})}{\partial x_{11}^2} = w_4 \frac{(a_0 + a_1 x_{21} + a_2 x_{22} + a_3 x_{21} x_{22} + a_4 x_{21}^2 + a_5 x_{22}^2)^{\frac{1}{2}}}{1.6 \cdot x_{22}} \cdot \frac{x_{21}}{2.7 \cdot (b_0 + b_1 x_{11} + b_2 x_{11}^2)} \quad (12)$$

$$y_5 = \frac{\partial^2 f(x_{21}, x_{22})}{\partial x_{11} \partial x_{12}} = w_5 \frac{(a_0 + a_1 x_{21} + a_2 x_{22} + a_3 x_{21} x_{22} + a_4 x_{21}^2 + a_5 x_{22}^2)^{\frac{1}{2}}}{1.6 \cdot x_{22}} \cdot \frac{x_{21}}{2.3 \cdot (b_0 + b_1 x_{11} + b_2 x_{12} + b_3 x_{11} x_{12})} \quad (13)$$

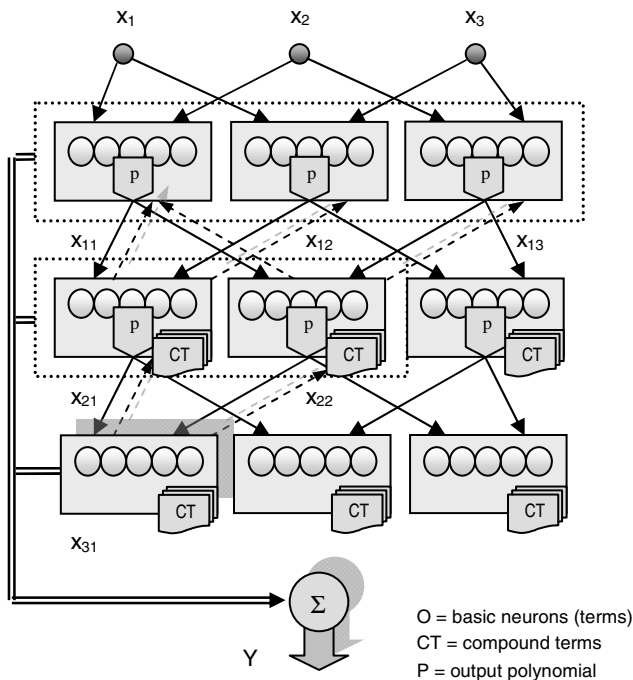


Fig. 3. 3-variable multi-layer D-PNN with 2-variable combination blocks

$$Y = \frac{\sum_{i=1}^k y_i}{k} \quad k = \text{actual number of active neurons} \quad (14)$$

Only some of all the potential combination DE terms (neurons) may participate in the DE composition, in despite of they have an adjustable term weight (w_i). A proper neuron combination, which substitutes for a DE solution, is not able to accept the disturbing effect of the rest of the neurons on the optimization of the parameters. D-PNN's total output Y is the arithmetic mean of all the active neuron output values so as to prevent a changeable number of neurons (of a combination) from influencing the total network output value (13). The selection of a fit neuron combination is a principal phase of the DE composition, performed simultaneously with polynomial parameter adjustment. It may apply the simulated annealing (SA) method, which employs a random search which not only accepts changes that decrease the objective function (assuming a minimization problem), but also some changes that increase it [8]. The error function, calculated using the root mean square error (RMSE) method (14), requires a minimization with respect to the polynomial parameters. It could be performed by means of the gradient steepest descent (GSD) method [1] supplied with sufficient random mutations to prevent from to be trapped to a local error depression.

$$E = \sqrt{\frac{\sum_{i=1}^M (y_i^d - y_i)^2}{M}} \rightarrow \min \quad (15)$$

$y_i^d = \text{desired output}$ $y_i = \text{estimated output for } i^{\text{th}} \text{ training vector}$

5 Power Demand Forecast Model Experiments

The presented D-PNN (Fig.3.) applied only 1-parametric hourly historical power demand time-series, free on-line available [11], to form a following 24-hour prediction.

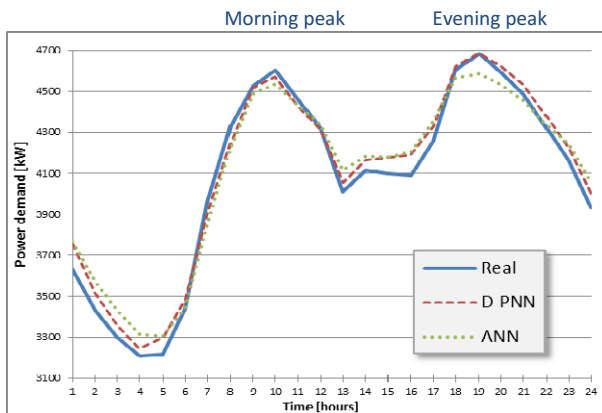


Fig. 4. TEPCO, 31.1.2012, Tuesday (D-PNN_{MAPE} = 1.37, ANN_{MAPE} = 2.01)

A power demand daily period consists typically of 2 peak-hours, morning (midday) and evening peak consumptions (Fig.4.). 3 consequent day power hourly values of previous week(s) of the same time points form the network input vector, a following day corresponding time variable defines the desired output. After training the network can predict the following day hour interval time-series with respect to the current week 3 past day power demand variables in a uniform time, as described in Chap.2. (Fig.4.- Fig.6.). Most models can obtain a very good approximation (Fig.4.), some are less accurate (Fig.6.) or even difficult to be formed. The mean absolute percentage error (MAPE) is a measure of accuracy of a method estimating time-series values.

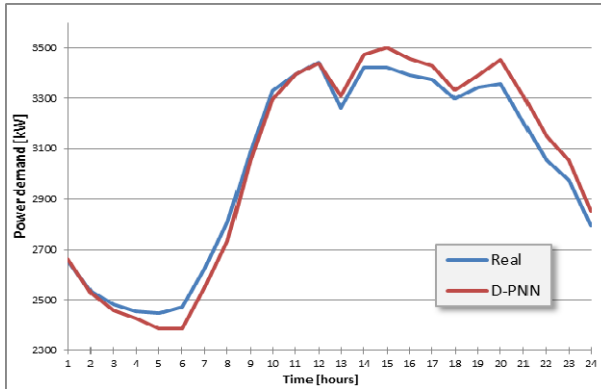


Fig. 5. TEPCO, 18.5.2012, Friday (MAPE = 1.74)

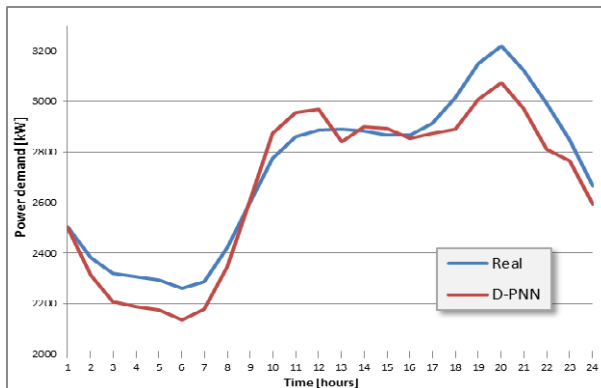


Fig. 6. TEPCO, 22.7.2012, Sunday (MAPE = 3.16)

Sundays, Saturdays and legal holidays power models result largely in higher inaccuracies, induced likely by more variable demands of weekends and holidays. The 2nd data set [12] experiments get with any worst results (Fig.7. and Fig.8.) than the 1st one [11]. Some week period models can succeed (Fig.7.), some embody high inaccuracies (Fig.8.), which unstable weather conditions can likely more influence. ANN predictions using the same method as D-PNN, get any worst results in the majority of cases.

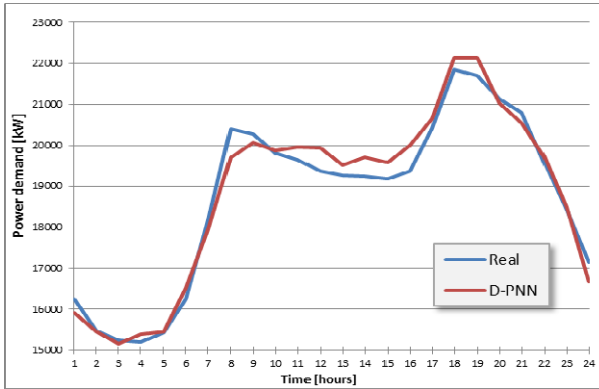


Fig. 7. IESO, 25.1.2012, Wednesday (MAPE = 1.47)

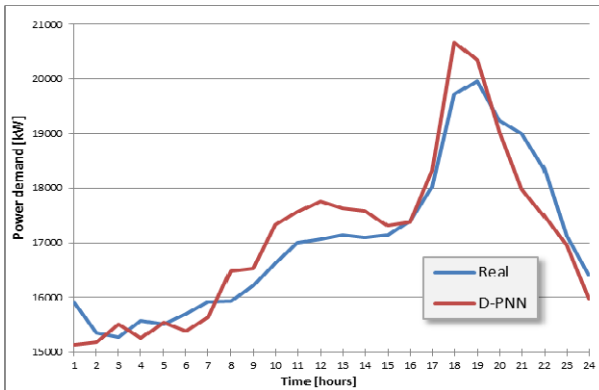


Fig. 8. IESO, 29.1.2012, Sunday (MAPE = 2.51)

The real model could be backward tested for corresponding days of last not trained week. 3-week training period seems to define an optimal learning scheme, it is better to apply less or more week-cycles (2-6) in some cases. The valid number of training week periods might be estimated in practice with respect to previous day(s) best testing results, as the progress of the subsequent daily cycles is quite of a uniform character. There is necessary to consider also legal holidays, which corresponding week Sundays (or working days) might replace in the learning scheme.

6 Conclusion

The presented method of daily power demand predictions is based on similarity data relations of subsequent day progress periods at the same time points of past week cycles. The power history affinities of past weeks daily periods, foregoing to the current week next day denomination, primarily influence the model solution accuracy, as no other considerable effects were taken into account, especially weather conditions.

The optimal prediction model could apply the power time-series combined with some input exogenous factors. Several previous week 3-day hourly series form a learning scheme of the neural network model, which after training apply as inputs the current week last 3-day power series to estimate following 24-hour values. The study applied 3 input vector variables using a new neural network type called the D-PNN, which results from the GMDH polynomial neural network. 1-parametric time-series function of data observations is described by an ordinary differential equation, which is formed and substituted by producing sum series of single and compound polynomial fractional derivative terms. D-PNN's relative non-linear data regression is contrary to the common soft-computing method approach, which applications are subjected to a fixed interval of absolute values.

Acknowledgement. The article has been elaborated in the framework of the IT4Innovations Centre of Excellence project, reg. no. CZ.1.05/1.1.00/02.0070 funded by Structural Funds of the European Union and state budget of the Czech Republic and in the framework of the project Opportunity for young researchers, reg. no. CZ.1.07/2.3.00/30.0016, supported by Operational Programme Education for Competitiveness and co-financed by the European Social Fund and the state budget of the Czech Republic.

References

1. Bertsimas, D., Tsitsiklis, J.: Simulated annealing. *Statistical Science* 8(1), 10–15 (1993)
2. Chan, K., Chau, W.Y.: Mathematical theory of reduction of physical parameters and similarity analysis. *International Journal of Theoretical Physics* 18, 835–844 (1979)
3. Darbellay, G.A., Slama, M.: Forecasting the short-term demand for electricity. *International Journal of Forecasting* 16, 71–83 (2000)
4. Garcia-Ascanio, K., Mate, C.: Electric power demand forecasting using interval time series. *Energy Policy* 38, 715–725 (2010)
5. Hippert, H.S., Pedreira, C.E., Souza, R.C.: Neural Networks for Short-Term Load Forecasting: A Review and Evaluation. *IEEE Transactions on Power Systems* 16(1) (2001)
6. Mamlook, R., Badran, O., Abdulhadi, E.: A fuzzy inference model for short-term load forecasting. *Energy Policy* 37, 1239–1248 (2009)
7. Nikolaev, N.Y., Iba, H.: *Adaptive Learning of Polynomial Networks*. Springer (2006)
8. Nikolaev, N.Y., Iba, H.: Polynomial harmonic GMDH learning networks for time series modelling. *Neural Networks* 16, 1527–1540 (2003)
9. Unsihuay-Vila, C., Zambroni, A.C., Marangon-Lima, J.W., Balestrassi, P.P.: Electricity demand and spot price forecasting using evolutionary computation combined with chaotic nonlinear dynamic model. *Electrical Power and Energy Systems* 32, 108–116 (2010)
10. Zjavka, L.: Recognition of Generalized Patterns by a Differential Polynomial Neural Network. *Engineering, Technology & Applied Science Research* 2(1) (2012)
11. TEPCO Tokyo Electric Power Company – past electricity demand data, <http://www.tepco.co.jp/en/forecast/html/download-e.html>
12. IESO Independent Electricity System Operator, Power to Ontario, Market data – Hourly demands 2002-2012, <http://www.ieso.ca/imoweb/marketdata/marketData.asp>

Design of Embedded Ethernet Interface Based on ARM11 and Implementation of Data Encryption*

Chunlei Fan, Zhiqiang Li, Qun Ding, and Songyan Liu**

Heilongjiang University,
Key Laboratory of Electronic Engineering College of Heilongjiang Province,
Harbin, China
liusongyan@hlju.edu.cn

Abstract. Network interface is studied based on S3C6410 processor and DM9000 Ethernet controller, the hardware circuit of network card interface is designed and developed with conciseness and stability. The DM9000 platform device drivers is designed in the aspect of software, then the development process of network card driver programs based on Linux operating system are introduced. The network card drivers are analyzed such as the initialization of network device, the process of sending and receiving data packet, etc. Finally, through to network data encryption of socket programs based on RC4 algorithm, this has confirmed the system accuracy and finished encryption of network datas.

Keywords: DM9000, network device driver, ARM11, Ethernet interface, network data encryption.

1 Introduction

As we enter the 21st century, with the development of the science and technology, the Internet is developing rapidly. Network development brings enormous convenience for modern society. It also has brought an unprecedented impact. At the same time, embedded system is more and more widely used. Embedded devices has been developed rapidly in the research lab and used widely in many fields such as industry, military department and personal consumption. Therefore, peoples requirement has increasingly expanding for system with functions of network access equipment. For today's embedded electronics information sharing is of great significance. This is of great importance in information sharing of electronic products for the embedded system [1].

Based on studying Linux embedded system and performance of hardware chip in detail, deciding to adopt S3C6410 high-speed processor and DM9000

* This paper was supported by Innovated Team Project of 'Modern Sensing Technology' in colleges and universities of Heilongjiang Province (No. 2012TD007) and Institutions of Higher Learning by the Specialized Research Fund for the Doctoral Degree (No. 20132301110004).

** Corresponding author.

chip to design and implement an embedded Ethernet access system. Besides, network data encryption of socket programs based on RC4 algorithm is designed to confirm system accuracy. The design will play a positive role in the research and application in the field of embedded Ethernet.

2 Designed of Hardware Interface

2.1 The Selection of Chip

In the aspect of hardware's choice for embedded system, the S3C6410 processor of Samsung Company is used. Compared with the previous S3C2410, S3C6410 is a 16/32-bit RISC microprocessor, which is designed to provide a cost-effective, low-power capabilities, high performance application processor. The S3C6410 has an optimized interface to external memory. These port support NOR Flash, Nand Flash type external memory and mobile DDR. To reduce total system cost and enhance overall functionality, S3C6410 includes many hardware peripherals such as 4-channel UART, 32-channel DMA, etc.

In addition, DM9000 is selected as Ethernet controller chip. The DM9000 is a fully integrated and cost-effective low pin count single chip fast Ethernet controller with a general processor interface, a 10/100M PHY and 4K double word SRAM. The PHY of the DM9000 can interface to the UTP3, 4, 5 in 10Base-T and UTP5 in 100Base-TX with HP Auto-MDIX [2]. It is fully compliant with the IEEE 802.3u Spec, etc. The internal block diagram of DM9000 is shown in figure 1.

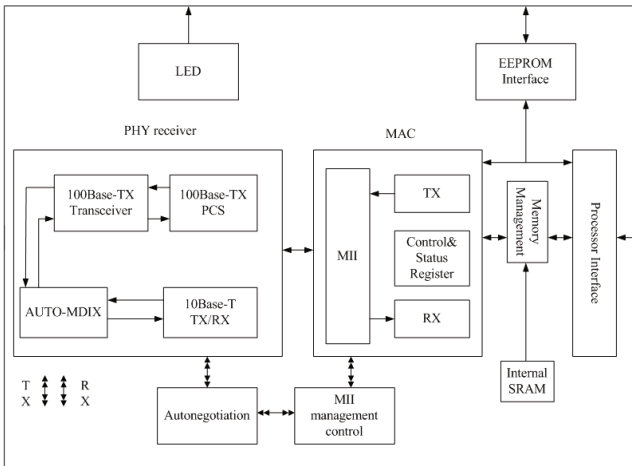


Fig. 1. The internal block diagram of DM9000

2.2 Overall Framework of Hardware Design

The design concepts of embedded system network interface use S3C6410 processor as the core of the system. It connects with Nand Flash and SDRAM so that Linux embedded operating system can run normally. In addition, a DM9000 independent module is designed for connect to the ARM11 processor by connectors. Besides, this system has serial port and JTAG port for program download and testing. The overall structure of hardware circuit is shown in figure 2.

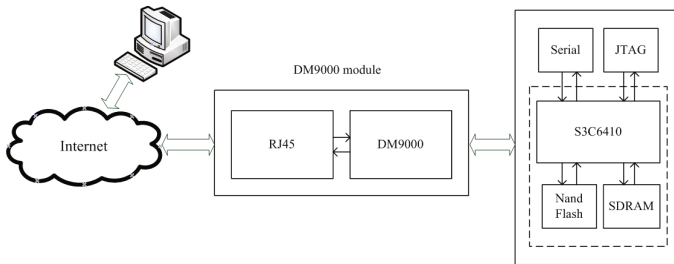


Fig. 2. Overall structure of hardware circuit

2.3 Main Circuit Design of Ethernet Interface

The interconnection design of S3C6410 and DM9000 are the most important in the hardware design of the embedded Ethernet interface. Ethernet micro-controller is connected to the bus of processor. Therefore, network data can be exchanged on the external bus. The wiring connection diagram is shown in figure 3.

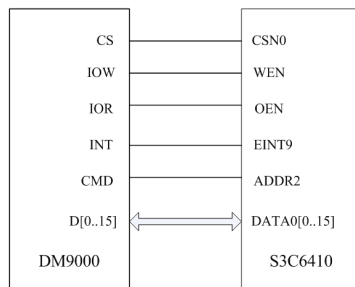


Fig. 3. The wiring diagram of S3C6410 and DM9000

The DM9000 of system adopts 16-bit working pattern. DM9000 data bus D[0..15] are connected to the processor's DATA[0..15] for network data transmission. Break request signal and EINT9 are linked together. The IOR and IOW serve as read/write command pin with low-level effectively. CS signal pin of chip select is connected to the processor's CSN0 and 0x10000000 as NIC

port address [3,4]. Therefore, DM9000 address port 0x10000000 and data port 0x10000004 are defined according to chip select CSN0. The access control of DM9000 is controlled by CMD pin. CMD pin is read as high-level for access to the data port and low-level access address port. Moreover, the input of address port is the data port register address before accessing any card registers. The address of the register should be saved in the address port [5].

3 Designed of DM9000 Driver Programs

3.1 Linux Network Driver Framework

Linux networking subsystem is divided into hardware layer, device driver layer, network protocol layer and application layer. Network protocol layer receives some network data packets, which are sent to the specific equipment by the related function of device drivers. Receiving function of device driver programs will do corresponding analysis to network data that are transmitted by communications equipment. At the same time, it is assembled into the corresponding data packets for the sake of network protocol layer. Therefore, the main work of a network device driver is providing services to its upper layer according to the specific hardware devices [6].

The diversity of network equipment is simplified by Linux system in order to provide the uniform and transparent interface for the user. All network hardware will be accessed through this interface. It provides a unified set of operations to all kinds of hardwares. This network interface can be regarded as sending and receiving data packets entity. This interface is net_device structure on Linux operating system. It defines some standard method for system's access and invoking of protocol layer, including the initialization of the init function, the open and stop function, hard_start_xmit function of processing packets and interrupt handling, etc. Network data packet is encapsulated by the sk_buff structure body. The working principle of the network device driver is shown in figure 4.

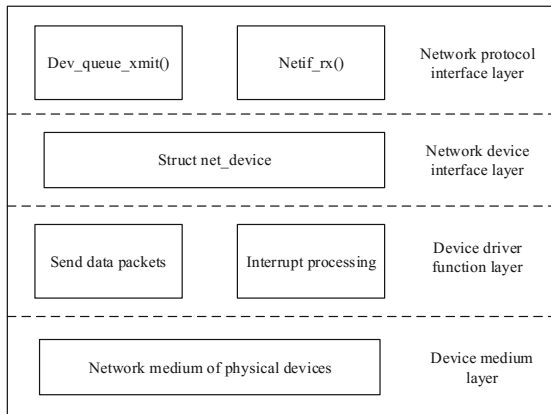


Fig. 4. Linux network driver architecture

3.2 Definition and Registration of DM9000 Platform Device

The DM9000 network device driver of this system is designed based on the platform driver architecture. Platform device as a virtual device can effectively simplify the design difficulty of the driver. It has two parts of `platform_device` and `platform_driver` [7]. The actual design sequence is shown in figure 5.

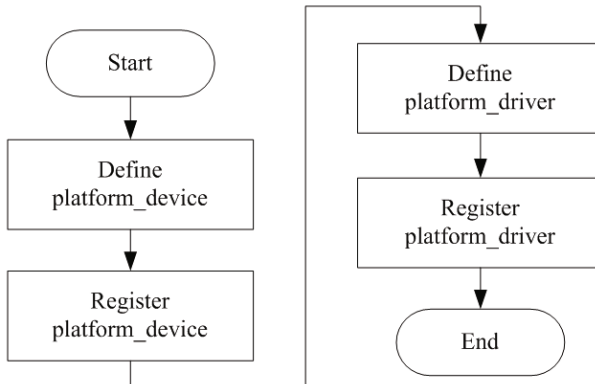


Fig. 5. Design flow chart of platform device drivers

Firstly, `s3c_device_dm9000` structure is defined as platform device according to design sequence of platform device drivers. The code to do this is shown below:

```

static struct platform_device s3c_device_dm9000 = {
    .name = "dm9000",
    .id = 0,
    .resource = dm9000_resources
}
  
```

Static structure resource allocates some resources for requirement of DM9000 network device, such as interrupt signal, chip selects, etc. The main code is as follows:

```

static struct resource dm9000_resources[] = {
    [0] = {
        .start = s3c64xx_pa_dm9000,
        .end = s3c64xx_pa_dm9000 + 3,
        .flags = IORESOURCE_MEM,
    },
    [2] = {
        .start = irq_eint(9),
        .end = irq_eint(9),
        .flags = IORESOURCE_IRQ | IRQF_TRIGGER_HIGH,
    },
}
}
  
```

Next, program need to finish platform_device structure's registration with code as follows:

```
static struct platform_device *smdk6410_devices[] __initdata={
#ifdef config_dm9000
&s3c_device_dm9000,
#endif
}
```

3.3 Crucial Function Analysis of DM9000 Platform Device

With those tasks complete, platform driver dm9000.c is written. The key functions of programs are analyzed and designed in detail in the thesis.

Platform driver trigger initialization process by defining dm9000_init function. The main function of platform_driver_register structure registers platform driver to Linux system. Besides, driver is added to bus so that driver can match with network device. At the same time, probe function is executed to detect equipment information and save the resources, according to information apply for memory and interrupts.

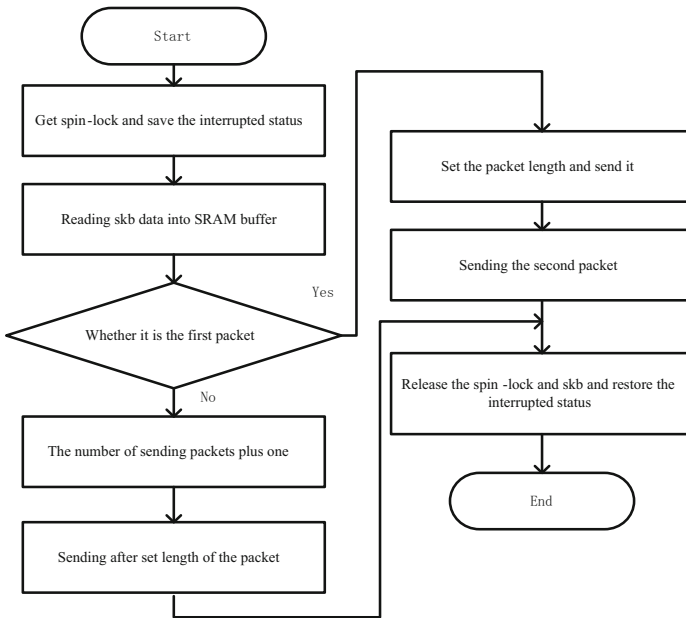


Fig. 6. The flow chart of sending data

Furthermore, sending and receiving of data packet is extremely important to network devices. Sending process of data packets can be described as follow. Driver activate DM9000 card and call the open function of net_device structure open

the network device when a network data needs to be sent. Then `hard_start_xmit` function is invoked by `dev_queue_xmit` structure so that the data of `sk_buff` structure can be sent to the network physical devices. DM9000 has 16 KB SRAM to act as sending and receiving data buffer. The address of data buffer ranges from 0x0000 to 0x0BFF. The area can save two complete Ethernet frame. The data frame length is written to the register TXPLH and TXPLL. The `sk_buff` will be released and returns zero value when packets are sent successfully. The driver will generate interrupt and inform system to send the next frame while first data completes transmission [8]. Flow chart of sending data is shown in figure 6.

DM9000 network equipment is through the interrupt mechanism to receive packets. Driver will generate an interrupt after network device receives a packet. `Sk_buff` structure is allocated in interrupt function for the sake of saving data. The address of receiving data buffer ranges from 0x0C00 to 0x3FFF, then the relevant information of data is filled to `sk_buff` after system read the data from a hardware in the receive buffer. Finally, `netif_rx` function is invoked to send the packet to upper network protocol layer. It means that system will generate an interrupt after network card receives a packet. In addition, processor read the data through MRCMDX or MRCMD register and the receiving packet format and function of each field as shown in table 1.

Table 1. Field contents of receiving data

Field type	Field description
Receiving data flag byte	01H for a data, 00H for no data
Status flag byte	Indicates that the data type
The length of the low byte	Receive data of low length identifier
The length of the high byte	Receive data of high length identifier
The data fields	Store the received data

4 The Test of DM9000 Network Driver

In order to verify the feasibility of the design scheme of Ethernet interface, we adopt to Linux network application programs based on the TCP stream socket. Therefore, a `client.c` and `server.c` are written by establishing the socket network to achieve the sending packets of client and server. The paper will regard the host as the server side and the target board as the client. RC4 encryption algorithm will embedded into the server program for the sake of encrypting network data. Further, to illustrate the correctness of the design scheme of Ethernet interface and guaranteeing the security of network data.

4.1 The Principle of RC4 Algorithm

RC4 algorithm is a stream ciphers with the block of length n . The internal state of algorithm contains $N = 2^n$ bytes of S box. RC4 algorithm consists of

two parts: key scheduling algorithm and pseudo-random generation algorithm. For one thing, random key K (typical length is 64 or 128 bits) produces an element $\{0, 1, \dots, N-1\}$ to form the initial arrangement of $S\{0, 1, \dots, N-1\}$. For another, PRGA generate pseudo-random key sequence $z[i]$ by initial arrangement S . Finally, $z[i]$ xor with plaintext to generate ciphertext [9,10,11].

4.2 The Test of Socket Programs

Firstly, cross development environment need to be established for embedded application. The design of the system installs Linux operating system of ubuntu9.10 version and cross-compilation tool chain of cross-4.2.2-eabi. DM9000 network drivers are modified in the source code after set up environment. Next, the kernel will be compiled by using the make command after configuration. Finally, compiling DM9000 driver into the kernel image file and generate zImage. The zImage, uboot, root file system burn to target board after completing the above steps [12].

We execute the client.c and server.c program after start Linux embedded system. Client side sends “socket-encryption-test” to server side. The data is encrypted based on RC4 algorithm before server side receives the data. We enter “socket-encryption-test” character via the keyboard on the client side. The data is encrypted based on RC4 algorithm before client side sends the data. The end result is shown below in figure 7. Then the data is decrypted by the some secret key after server side receives the ciphertext. The ciphertext and plaintext are printed on the server side. The end result is shown below in figure and 8.



Fig. 7. The display of client program

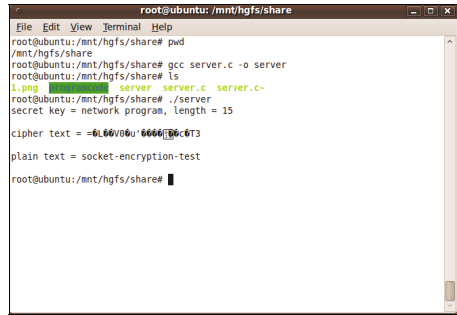


Fig. 8. The display of server program

5 Conclusion

This paper introduces a hardware design of embedded Ethernet interface based on S3C6410 ARM11 processor and DM9000 Ethernet controller. It is very simple to implement, extend and renew this scheme, which especially fits the Internet access and security of embedded products. In recent years, due to the phenomenal

growth of networking and communication systems, embedded system becomes increasingly important on the Internet and other technical areas. Therefore, the works of thesis supply the future development of system with theory support and hardware environment. There are also important research and reference values in embedded network security device.

References

1. Fang, A., Zhang, Y.: Design of High-Speed Electric Power Network Data Acquisition System Based on ARM and Embedded Ethernet. *J. Adv. Intell. Comput.* 139, 455–459 (2012)
2. Zhao, R.M., Wang, M.: Design of ARM-based embedded Ethernet interface. *J. Comput. Eng. Technol.* 4, 268–270 (2010)
3. Ge, Y.M.: Ethernet interface embedded system design. *J. Electron. Technol. Appl.* 7, 25–27 (2002)
4. Zhang, J.: Embedded Ethernet technology and its application in the field of industrial measurement and controlment. *J. Instrum. Technol. Sens.* 5, 36–37 (2003)
5. Chen, S.K., Jinag, S.J.: Design of Embedded Network Interface Controller Based on ARM9 and ARMLinux. *J. Commun. Comput. Inf. Sci.* 34, 142–149 (2009)
6. Matsui, M.: Key Collisions of the RC4 Stream Cipher. In: Dunkelmann, O. (ed.) *FSE 2009. LNCS*, vol. 5665, pp. 38–50. Springer, Heidelberg (2009)
7. Attacks on the RC4 Stream Cipher,
[http://cage.agent.be/\\$\sim\\$klein/RC4/RC4-en](http://cage.agent.be/\simklein/RC4/RC4-en)
8. Yuan, W.Q.: Research and Implementation of Embedded Interface of Ethernet. *J. Instrum. Technol. Sens.* 11, 59–61 (2008)
9. Li, J.X., Zhang, C.: Design and Implementation of Network Card Interface Based on ARM and DM9000. *J. Comput. Inf. Technol.* 24, 123–125 (2008)
10. Zhou, J.Q.: Development of Network Device Driver Based on Linux. *J. Emb. Syst. Eng.* 30, 5124–5127 (2009)
11. Yang, R., Cai, H.: Research and implement of Ethernet interface based on embedded system. *J. Comput. Intell. Des.* 2, 288–291 (2009)
12. Fan, Z.: Fundamentals of network security. *J. Comput. Commun.* 17, 121–136 (2009)

The Evaluation of the Business Operation Performance by Applying Grey Relational Analysis

Dingtao Zhao¹, Su-Hui Kuo¹, and Tien-Chin Wang^{2,*}

¹University of Science and Technology of China, China
box@ustc.edu.cn, kcsophia@yahoo.com.tw

²Dept. of International Business, National Kaohsiung University of Applied Sciences,
415 Chien-Kung Road, Kaohsiung City 80778, Taiwan
tcwang@kuas.edu.tw

Abstract. The effect of the business operation performance has a great influence on the growth and development of a corporation; therefore, the purpose of measuring the business operation performance is to understand whether the application and allocation of the resources in a corporation have reach the optimality and the completeness progress of goals, and these all provide the management with the essential information as valuable references for possible correction plans or policy decisions with a view to enhancing business competitiveness. Furthermore, investors are able to develop judgment of investment according to the results of evaluating the business operation performance. Takes 9 tourist hotels in Taiwan as study objects, collect related financial data in 2012 from Taiwan Stock Exchange (TWSE) and designate 6 financial ratios—Current Ratio, Fixed assets turnover ratio, Debt Ratio, Return on Equity (ROE), Growth Rate of Operating Income and Account Receivable Turnover Ratio—as evaluation indicators. By applying the method of Grey Relational Analysis, we obtained the grades of performance evaluation of the 9 objects and then arranged them in order. Afterwards, according to the ranking and through comparison among the evaluation results of the 9 corporations, we found the best corporate as a “Benchmark”, which is to be a model for other companies in the same industry and to serve as good reference for general inventors when making investment decisions.

Keywords: Performance Evaluation, Grey Relational Analysis, Benchmarking, Financial Indicators.

1 Introduction

Business operating performance is always being taken as an essential indicator of corporation development, for the management, through the evaluation of operating indicator, can fully realize the efficiency and effectiveness of operation of enterprise resources, then locating problems and adjusting operation policies accordingly, so as to gain more competitive advantages. From that reason, an enterprise must deliberately and objectively evaluate business operating performance of itself to be sure of

* Corresponding author.

future business growth and optimistic development and to reduce the risk of incorrect investment decisions. The public is always greatly concerned about an enterprise's financial performance and look upon it as an indicator of whether to invest, for this indicator is to help them select desirable investment targets with high profitability and growth dynamic and receive better returns in stock market.

This study took 9 tourist hotels as study objects and applied the method of Grey Relational Analysis to analyze their financial data, which are open resources and available to the public, and to evaluate business operating performances of each. With the performance evaluation results, not only is the public capable of having a clear idea about an enterprise's financial status and state of operation, but corporations also acquaint themselves with where they are in the industry by comparing relative operating performance with others and find out the best enterprise as the benchmark, a model to imitate and learn from.

According to the study purposes, the following is the content of the study: (1) Settle the sample data of the study—9 tourist hotels as the study objects. (2)select the financial-related indexes and acquire data needed from the open financial statements in 2012 from Taiwan Stock Exchange (TWSE). (3)Get an order list of business operating performance from the outputs calculated by applying Grey Relational Analysis. (4) Compare the ranking of business operating performance of each enterprise and analyze the differences among them to discover the best one to be the “benchmark”, which is to be the model of other corporations in the same filed or industry and to provide the general investors with the reference basis when making investment decision.

2 Literature Review

2.1 Evaluation of Business Operating Performance

The importance of performance in management is beyond doubt. Early scholars considered performance the measure of the completeness of goals in an organization, which consists of efficiency and effectiveness. Efficiency is to select optimal action plans and use correct methods to achieve designated aims; effectiveness is the rate of achievement of goals. In brief, the true meaning of performance in management lies in how to achieve the goals of an organization and enhance the effectiveness of it through the powers from people or groups.

However, nowadays, opinions are widely divided regarding the measure of performance. Some studies took financial indexes such as return on investment and profitability as the standards of performance evaluation, while others used qualitative data such as productivity, organization commitment and working satisfaction of employees to explain the business operating performance of a company. Venkatraman and Ramanujam (1986) [18] stated the measuring of business performance can be divided into three catalogues—financial performance indicator (the amount of sales growth), operating performance indicator and organizing performance indicator. Financial performance indicator is the most widely used concept. As for operating performance indicator, in addition to the concept of financial performance indicators, it involves

non-financial performance indicators such as market shares, product quality, introduction of new products and the creating of additional value. Considering the date and information from financial statements made by enterprises are more objective and less controversial, the study uses the financial indicators as the measure of business operating performance.

Scholars have performed many studies related to the evaluation of business operating performance, and they used different financial ratio variables, analytical method and study samples for difference study subjects. Yang et al. (1999) [20] took financial ratios as study variables to establish prediction model of financial crisis of enterprise. Feng and Wang (2000) [12] used financial ratios to evaluate the performance of aviation industry. From above reference documents, the analysis of financial ratios truly an effective tool to measure the operating performance of enterprises or to predict the possible failure of them. Financial statements not only provide the information for measuring operating performance but also contribute to planning current or future activities and even works well as a reference basis for future decisions.

2.2 Benchmarking

The concept of benchmarking is about finding out the outstanding corporation in an industry and regard it as a model to improve the process performance and competitive advantages by learning its goodness and strengths. American Productivity and Quality Center (APQC) (1993) [1] defined benchmarking as “a systematic and continuous process of measuring. By assessing practices of the best-in-class, determine the extent to which they might be adapted to achieve superior performance.”

From the above definitions, we knew benchmarking aims to improve running performance of enterprise. By accessing and implementing the best practice in the field, we derive action plans to improve our weakness and defects. As Boxwell (1994) [2] said, the main purpose of enterprises’ practicing benchmarking is to root out the problems in business running and seek for solutions. Codling (1998) [3] suggested that enterprises focus on the parts which can bring up the most benefits by practicing benchmarking.

3 Research Methods

3.1 The Selecting of Study Objects and the Evaluation Indexes of Performance

The study objects of this study are 9 tourist hotels in Taiwan, evaluated for operating performance through analyzing financial statements in 2012, which are open and available to the public in Taiwan Stock Exchange (TWSE). Seeing that financial ratios indicate the performance an enterprise practices during a period of time and also offer a better picture of what an enterprise’ health looks like. Thus, as Venkatraman and Ramanujam (1986) [18] conducted, this study also mainly uses financial ratios to analyze the data and information collected and takes them as measures of performance. In selecting indicators, after reviewing related documents and discussions

with scholars and experts in the field, This study probes into business operation performance from 6 significant financial ratios—solvent capability, asset management, financial structure, profitability, growth capability, activity capability, organized as Table 1 below.

Table 1. Indicators of performance evaluation

Dimensions of Performance	Measuring indicators	Definitions
Solvent capability	Current ratio	Current ratio=current asset / current liability
Asset management	Fixed assets turnover ratio	Fixed assets turnover ratio= net sales / average net fixed assets
Financial structure	Debt ratio	Debt ratio=Total liabilities / Total assets
Profitability	Return on Equity (ROE)	ROE=Net income after taxes / average equity
Growth capability	Operating income growth rate	Operating income growth rate = current operating income / former operating income
Activity capability	Account receivable turnover ratio	Account receivable turnover ratio =net sales / average account receivable

Sources: This research summarized

3.2 Principles of Grey Relational Analysis (GRA)

The Grey System theory was proposed by Deng in 1982 [8], and has proven useful for dealing with poor quality, incomplete, and uncertain data. A grey system is a system which contains insufficient information [14], and the theory is enormously popular due to its ability to work with these problems of uncertainty, multi-input, discrete data and not enough data. Grey relational analysis [9] is one of two mainstays for the grey system theory. The GRA can be used to capture the correlations among factors and candidates of a system. One of the advantages of GRA is that the quantitative and qualitative relationships can be identified from numerous factors with insufficient information. Many traditional mathematic correlations analyses can render not only coefficients among factors but also relevant levels [4]. In the following, we briefly review some definitions of grey relational analysis from [4,11,13,15]. These basic definitions and notations below will be used throughout the paper unless otherwise stated.

Definition 1. Let the original reference sequences and comparability sequences be noted as $x_0^{(0)}(k)$ and $x_i^{(0)}(k), i=1,2,\dots,n; k=1,2,\dots,m$, respectively. Here, the original reference sequence and comparability sequences can be expressed as Eq.(1) and Eq.(2):

$$x_0^{(0)} = \{x_0^{(0)}(k)\} = (x_0^{(0)}(1), \dots, x_0^{(0)}(m)) \in X, k = 1, 2, \dots, m \tag{1}$$

$$x_i^{(0)} = \{x_i^{(0)}(k)\} = (x_i^{(0)}(1), \dots, x_i^{(0)}(m)) \in X \tag{2}$$

$i = 1, 2, \dots, n; k = 1, 2, \dots, m$

Definition 2. If the target value in original sequence has the characteristic of “the-larger-the better”, the original sequence will be noted as follows: [19]

$$x_i^*(k) = \frac{x_i^{(0)}(k) - \min x_i^{(0)}(k)}{\max x_i^{(0)}(k) - \min x_i^{(0)}(k)} \tag{3}$$

When the value is “the-smaller-the better”, the original sequence can be denoted as follows:

$$x_i^*(k) = \frac{\max x_i^{(0)}(k) - x_i^{(0)}(k)}{\max x_i^{(0)}(k) - \min x_i^{(0)}(k)} \tag{4}$$

Rather, if the target value (*OB*) is specified between the maximum and the minimum, the original sequence is defined in this form:

$$x_i^*(k) = 1 - \frac{|x_i^{(0)}(k) - OB|}{\max\{\max[x_i^{(0)}(k)] - OB, \min[x_i^{(0)}(k)]\}} \tag{5}$$

Definition 3. The grey relational coefficient can be defined as follows:

$$\xi(x_i^{(0)}(k), x_j^{(0)}(k)) = \frac{\Delta_{\min.} + \rho\Delta_{\max.}}{\Delta_{ij}(k) + \rho\Delta_{\max.}} \tag{6}$$

$$\Delta_{\max.} = \max_{\forall_i} \max_{\forall_j} \|x_i^{(0)}(k) - x_j^{(0)}(k)\| \tag{7}$$

$$\Delta_{\min.} = \min_{\forall_i} \min_{\forall_j} \|x_i^{(0)}(k) - x_j^{(0)}(k)\|$$

The symbol ρ is a distinguish coefficient, taken between 0 and 1.0, frequently taken as 0.5.

Definition 4. The grey relational grade is taken as follows:

$$\gamma(x_i^{(0)}(k), x_j^{(0)}(k)) = \sum_{k=1}^m w_k \xi(x_i^{(0)}(k), x_j^{(0)}(k)), \tag{8}$$

$$\sum_{k=1}^m w_k = 1$$

The relational grades are numerical measures of the influence of factors on the objective values, and the numeric values are among 0 and 1. Generally, $\gamma > 0.9$ indicates a marked influence, $\gamma > 0.8$ a relatively marked influence, $\gamma > 0.7$ a noticeable influence, and $\gamma < 0.6$ a negligible influence. [5]

4 Data Analysis and Results

The study started the evaluation of business operation performance with 9 tourist hotels in Taiwan as study objects. To avoid possible trouble or incontinence, let (A_1, A_2, \dots, A_9) represent the 9 tourist hotels. Current ratio (C_1), Fixed assets turnover ratio (C_2), debt ratio (C_3), return on equity (C_4), operating income growth rate (C_5) and account receivable turnover ratio (C_6), the six financial indexes, are to be the evaluating criteria to measure operating performance. Table 2 shows the data and information of financial ratios extracted from the financial statements in 2012 given by the companies and all acquirable in Taiwan Stock Exchange (TWSE).

Systematic approaches of the computational procedures of grey relational analysis are described below:

Step 1. Initialize and transfer the original sequences data

Retrieved from the database, the initial original sequence (set out in Table 2) for these six criteria and 9 tourist hotels can be represented in matrix as:

$$X_0 = \begin{matrix} & C_1 & C_2 & \cdots & C_6 \\ \begin{matrix} A_1 \\ A_2 \\ \vdots \\ A_9 \end{matrix} & \begin{bmatrix} x_1(1) & x_1(2) & \cdots & x_1(6) \\ x_2(1) & x_2(2) & \cdots & x_2(6) \\ \vdots & \vdots & \vdots & \vdots \\ x_9(1) & x_9(2) & \cdots & x_9(6) \end{bmatrix} \end{matrix}$$

The symbol C_1, \dots, C_6 represents the six criteria of evaluation. The values of the original data must be normalized to have the same order, and are commonly normalized by mean value:

$$y_i = \frac{x_i(k)}{\frac{1}{m} \sum_{k=1}^m x_i(k)} \quad m = 9$$

Step 2. Determine the reference sequences and comparability sequences

Values of the reference sequences are the maximum in each criterion listed in Table 3. Reference sequences and comparability sequences are shown as Table 3.

Step 3. Determine the grey relational deviation sequences.

The deviation sequences Δ_{0i} can be defined as:

$$\Delta_{0i} = \left\| x_0^*(k) - x_i^*(k) \right\| \tag{9}$$

All the deviation values of comparability sequences processed by Eq. (9) can be seen in Table 4.

Step 4. Identify the maximum and minimum deviation.

Eq. (10) is adopted to perform the maximum and minimum deviation, Δ_{\max} is the absolute maximum deviation value between $x_0^{(0)}(k)$ and $x_i^{(0)}(k)$, and Δ_{\min} is the absolute minimum deviation value between $x_0^{(0)}(k)$ and $x_i^{(0)}(k)$.

$$\begin{aligned} \Delta_{\max} &= \max |x_0^{(0)}(k) - x_i^{(0)}(k)| \\ \Delta_{\min} &= \min |x_0^{(0)}(k) - x_i^{(0)}(k)| \end{aligned} \tag{10}$$

Investigating the values presented in Table 4, we find that $\Delta_{\max}(k)$ and $\Delta_{\min}(k)$ are taken as follows:

$$\begin{aligned} \Delta_{\max}(k) &= \Delta_{02}(5) = |-2.4204 - 10.6120| = 13.0324 \\ \Delta_{\min}(k) &= \Delta_{04}(1) = |5.9230 - 5.9230| = 0 \end{aligned}$$

Step 5. Produce the grey relational coefficient ξ .

The value of the distinguishing coefficient ρ lies between real numbers 0 and 1, that is, $\rho \in [0,1]$. This study taken it as 0.5.

Table 2. Original sequences data

	C_1	C_2	C_3	C_4	C_5	C_6
x_1	17.00	0.34	30.00	7.22	-4.69	36.00
x_2	63.41	0.36	51.96	-2.10	-12.48	37.51
x_3	159.51	0.42	24.90	3.34	2.27	25.36
x_4	2353.60	0.30	12.50	4.50	4.57	56.80
x_5	54.44	2.70	39.97	34.98	3.07	21.35
x_6	32.69	0.30	39.74	6.21	-6.67	25.78
x_7	427.75	0.53	24.37	5.32	3.67	17.81
x_8	396.56	4.26	11.00	12.40	1.95	103.34
x_9	71.35	0.81	47.63	6.38	54.71	10.48

Table 3. Normalization of original sequences data and comparability sequences

	C_1	C_2	C_3	C_4	C_5	C_6
x_0	5.9230	3.8263	0.3510	4.0233	10.6120	2.7810
x_1	0.0428	0.3054	0.9572	0.8304	-0.9090	0.9688
x_2	0.1596	0.3234	1.6579	-0.2415	-2.4204	1.0094
x_3	0.4014	0.3772	0.7945	0.3842	0.4399	0.6825
x_4	5.9230	0.2695	0.3988	0.5176	0.8859	1.5286
x_5	0.1370	2.4251	1.2753	4.0233	0.5950	0.5746
x_6	0.0823	0.2695	1.2680	0.7142	-1.2945	0.6938
x_7	1.0765	0.4760	0.7776	0.6119	0.7122	0.4793
x_8	0.9980	3.8263	0.3510	1.4262	0.3790	2.7810
x_9	0.1796	0.7275	1.5197	0.7338	10.6120	0.2820

Table 4. Grey relational deviation sequences

	C_1	C_2	C_3	C_4	C_5	C_6
Δ_1	5.8802	3.5210	0.6062	3.1928	11.5210	1.8122
Δ_2	5.7634	3.5030	1.3069	4.2648	13.0324	1.7716
Δ_3	5.5216	3.4491	0.4435	3.6391	10.1721	2.0986
Δ_4	0.0000	3.5569	0.0479	3.5057	9.7261	1.2525
Δ_5	5.7860	1.4012	0.9243	0.0000	10.0169	2.2065
Δ_6	5.8407	3.5569	0.9170	3.3090	11.9064	2.0873
Δ_7	4.8465	3.3503	0.4266	3.4114	9.8997	2.3017
Δ_8	4.9250	0.0000	0.0000	2.5971	10.2330	0.0000
Δ_9	5.7434	3.0988	1.1688	3.2895	0.0000	2.4990

Step 6. Grey relational grade for 9 tourist hotels

Applying the data in Table 4 to the following equation, and calculating grey relational grades of the 9 tourist hotels, where w_k is the weight of each criterion and we set it at 1/6 in this research.

$$\gamma(x_0^{(0)}(k), x_j^{(0)}(k)) = \sum_{k=1}^6 \left[\xi(x_0^{(0)}(k), x_j^{(0)}(k)) \times w_k \right]$$

From Table 5, we prioritize the 9 tourist hotels in accordance with their grey relational degrees. The larger the degree toward the $x_0^{(0)}(k)$ implies the better operation performance of hotel. Therefore, the ascend relational rank for tourist hotels are substituted as follows:

$$\gamma_{08} \succ \gamma_{04} \succ \gamma_{09} \succ \gamma_{05} \succ \gamma_{07} \succ \gamma_{03} \succ \gamma_{01} \succ \gamma_{02}$$

Table 5. Grey relational coefficient computation

	γ_{0i}	C_1	C_2	C_3	C_4	C_5	C_6	Rank
$\xi(x_0(k), x_1(k))$	0.6508	0.5257	0.6492	0.9149	0.6711	0.3613	0.7824	7
$\xi(x_0(k), x_2(k))$	0.6230	0.5307	0.6504	0.8329	0.6044	0.3333	0.7862	9
$\xi(x_0(k), x_3(k))$	0.6533	0.5413	0.6539	0.9363	0.6417	0.3905	0.7564	6
$\xi(x_0(k), x_4(k))$	0.7550	1.0000	0.6469	0.9927	0.6502	0.4012	0.8388	2
$\xi(x_0(k), x_5(k))$	0.7283	0.5297	0.8230	0.8758	1.0000	0.3941	0.7470	4
$\xi(x_0(k), x_6(k))$	0.6375	0.5273	0.6469	0.8766	0.6632	0.3537	0.7574	8
$\xi(x_0(k), x_7(k))$	0.6608	0.5735	0.6604	0.9386	0.6564	0.3969	0.7390	5
$\xi(x_0(k), x_8(k))$	0.7789	0.5695	1.0000	1.0000	0.7150	0.3890	1.0000	1
$\xi(x_0(k), x_9(k))$	0.7407	0.5315	0.6777	0.8479	0.6645	1.0000	0.7228	3

From Table 2, the ranking of performance should be $A_8 > A_4 > A_9 > A_5 > A_7 > A_3 > A_1 > A_2$. Per this study result, A_8 is the best practice of the **9 tourist hotels** in 2012. And the result also can be verified by the fact that the return on equity (C_4) and operating income growth rate of A_9 are both higher than the others, which indicates that the company's profitability is at the top and its operating competency has attained an acceptable level.

The performance result of A_4 is the second only to A_8 . Compared the performances of the two companies in each evaluative indicator from Table 2, the Current ratio (C_1) of A_4 is higher than A_8 and its fixed assets turnover ratio lower, which shows A_4 has been at a certain level in dimensions of solvent capability, while the other indicators all shows A_4 is inferior than A_8 . For the above reasons, it is recommended that the company A_4 make the most of benchmarking to foster the advantages of itself, improve the disadvantages and eventually promote and advance the operating performance of the company.

Moreover, we found the performance result of A_9 is next best to A_4 . Compared the performances of the two companies in each evaluative indicator from Table 2, the return on equity (C_4), operating income growth rate (C_5) of A_9 is higher than A_4 , but its debt ratio is obviously higher than A_4 . This indicates the effectiveness of production and sales of A_9 is feeble and results in less effective in business operating. Also, the great extent of debt is unfavorable to creditors. For the above reasons, the management of the company A_9 is recommended to practice benchmarking to improve the business operating performance.

5 Conclusion

This study is aimed at the 9 tourist hotels in Taiwan and utilizes financial indexes from open resources to analyze; with the application of the method of Grey Relational Analysis, we obtain the grades of performance evaluation and rank them accordingly. The following are conclusions derived from empirical analysis of the study: This study selects appropriate financial indicators to measure operating performance and each corresponds to a certain performance dimension, which gives the measuring more comprehensiveness integrity. The study provides a way to evaluate business operating performance and assists business administration to have a good grasp of the performance discrepancy in the industry. Making good use of benchmarking encourages business competitiveness and benefit operating performance by mutual learning and imitating, adjusting business policies, appropriately allocating business resources and searching out the main factors in such a difference. Therefore, the evaluation results of business operating performance is valuable and of considerable referential importance for both the management of a corporation and the general investors.

References

1. American Productivity & Quality Center (APQC): *The Benchmarking Management Guide*. Productivity Press, Portland (1993)
2. Boxwell, R.J.: *Benchmarking for Competitive Advantage*. McGraw-Hill, NY (1994)
3. Codling, S.: *Benchmarking*. Gower Publishing, Great Britain (1998)
4. Chang, T.C., Lin, S.J.: Grey relation analysis of carbon dioxide emission from industrial production and energy uses in Taiwan. *Journal of Environmental Management* 56, 247–257 (1999)
5. Fu, C., Zheng, J., Zhao, J., Xu, W.: Application of grey relational analysis for corrosion failure of oil tubes. *Corrosion Science* 43, 881–889 (2001)
6. Chen, S.T., Hwang, C.L.: *Fuzzy multiple attribute decision making methods and applications*, Germany: Springer-Verlag, pp. 1–5. Springer, Germany (1992)
7. Codling, S.: *Benchmarking*. Gower Publishing, Great Britain (1998)
8. Deng, J.L.: *The Essential Methods of Grey Systems*. Huazhong University of Science and Technology Press, Wuhan (1992)
9. Deng, J.L.: *The Course on Grey System Theory*, pp. 91–92. Huazhong University of Science & Technology Publish House, Wuhan (1990)
10. Deng, J.L.: Introduction to grey system theory. *The Journal of Grey System* 1(1), 1–24 (1989)
11. Deng, J.L.: Control problems of grey system. *Systems Control Letters* 1(5), 288–294 (1982)
12. Feng, C.-M., Wang, R.-T.: Performance evaluation for airlines including the consideration of financial ratios. *Journal of Air Transport Management* 6, 133–142 (2000)
13. Fu, C.Y., Zheng, J.S., Zhao, J.M., Xu, W.D.: Application of grey relational analysis for corrosion failure of oil tubes. *Corrosion Science* 43, 881–889 (2001)
14. Hu, Y.C., Chen, R.S., Hsu, Y.T., Tzeng, G.H.: Grey self-organizing feature maps. *Neurocomputing* 48, 863–877 (2002)
15. Huang, Y.P., Huang, C.C.: The integration and application of fuzzy and grey modeling methods. *Fuzzy Sets and Systems* 78(1), 107–119 (1996)
16. Kao, P.S., Hochen, H.: Optimization of electrochemical polishing of stainless steel by grey relational analysis. *Journal of Materials Processing Technology* 140, 255–259 (2003)
17. Liu, T.Y., Yeh, J., Chen, C.M., Hsu, Y.T.: A GreyART system for grey information processing. *Neurocomputing* 56, 407–414 (2004)
18. Venkatraman, N., Ramanunjam, V.: Measurement of business performance on strategy research: a comparison of approach. *Academy of Management Review* 11(4), 801–814 (1986)
19. Wu, J.H., Chen, C.B.: An alternative form for grey relational grades. *Journal of Grey System* 11(1), 7–12 (1999)
20. Yang, Z.R., Platt, M.B., Platt, H.D.: Probabilistic neural networks in bankruptcy prediction. *Journal of Business Research* 44(2), 67–74 (1999)

An Echo-Aided Bat Algorithm to Construct Topology of Spanning Tree in Wireless Sensor Networks

Yi-Ting Chen, Ming-Te Tsai, Bin-Yih Liao, Jeng-Shyang Pan, and Mong-Fong Horng

Department of Electronics Engineering,
National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan
{ytchen,mttsai byliao,jspan,mfhorng}@bit.kuas.edu.tw

Abstract. Echo-Aided Bat Algorithm is proved a good evolutionary computing to solve continuous problems in previous investigate. In this study, EABA is applied to solve the discrete problem that construct a network topology with spanning tree in wireless sensor networks (WSNs). In this application, the presentation of bat and design of fitness function perhaps affect the evolutionary results. For the demonstration of simulated results, EABA still presents a satisfied performance in some scenarios designed in this study. In addition, the constructed network topology of spanning tree based on global optimum solution found by EABA is used to estimate the network lifetime. Overall, this framework including network topology constructed by EABD and network lifetime estimated by transmission simulation is created in this study to aid the plan for duration of WSN deployment.

Keywords: Echo-Aided Bat Algorithm, Spanning Tree Network Topology, Wireless Sensor Network.

1 Introduction

The design of network topology affects the performances of WSNs. In WSNs, the sensors are unable to recharge. Once the energy of one sensor is exhausted, the performances of network will be impacted due to isolated node or broken data route. On the other hands, the over connections cause serious interference and power consumption. Therefore, minimum spanning tree is the perfect network topology [1-3]. The spanning tree can be defined as follows: given a undirected graph $G=(V, E)$ and edge cost $c_{i,j}$ for $\{i,j\} \in E$, select a set of edges with minimum cost that connects all of the nodes. Furthermore, find a spanning tree with minimum cost is a NP-hard problem [4,5]. In this study, an echo-aide bat algorithm (EABA) [6] is suggested to solve this NP-hard problem. The performance of EABA is substantially improved based on original bat algorithm (BA) [7,8] in continuous NP-hard problems. In order to prove that EABA is an excellent evolutionary computing in continuous and discrete problems, EABA is applied in an application that construct a network topology with spanning tree in WSN to evaluate the performance of algorithm in this study. In addition, the constructed network topology with spanning tree is adopted to estimate the network lifetime. As a whole, this designed framework including evolutionary

computing and network lifetime simulation benefits the performance evaluation of network before WSN deployment.

The rest of this study is organized as follows: the detail operation of EABA is described in Section 2. Next, the application scenario and EABA how to construct a network topology with spanning tree are illustrated in Section 3. In Section 4, a series of simulation is designed to verify the performance of EABA in discrete problem and estimate the lifetime of constructed network topology. Finally, the conclusion and future work are outlined in Section 5.

2 Related Works

2.1 An Echo-Aided Bat Algorithm

Bat Algorithm (BA) is a nature-inspired algorithm and proposed by Xin-She Yang to solve the optimization problems of single objective and multi-objectives [49, 50]. All bats have an ability to sense distance and know the difference between prey and background barriers. This ability is called echolocation. Bats randomly fly by velocity (v_i) with a fixed frequency (f_{min}), varying wavelength (λ), adjustable pulse emission rate (r_i) and changeable loudness (A_0) at position (x_i) to search the prey. The frequency f_i in bat i be assumed from f_{min} to f_{max} . The loudness is assumed between 1 and 2 as well as varies from a large (positive) A_0 to a minimum constant value A_{min} . The pulse emission rate is set between 0 and 1. In the evolution procedure of bat algorithm, the frequency is used to adjust the velocity. But this frequency is determined randomly with distribution uniform. And this adjusted velocity will affect the updated position of bat. In this manner, there are some errors when the bats search new position.

For this issue, an echo-aided bat algorithm supporting measurable movement is proposed in [6]. This measured echo time between bat and objective is utilized to modify the velocity of bat. The round trip time from bat to objective can estimate the distance from bat to objective. If the echo time is longer to indicate that the bat away from the objective, the bat should increase the velocity to approach to the objective. On the contrary, when the bat approaches to the objective, the echo time is short. The bat should decrease the velocity to slightly search the better solution near the objective. This innovative conception attempts to modify the error caused by frequency with random number to improve the algorithm performance. The detailed operation of EABA is illustrated as follows:

- Step 1. Initialization: In this step, there are many parameters to be assumed including iteration ($iter$), population size (PS), quality ratio (r_q) and dimension (d) of search space for algorithm. Then, for the property of bats, the frequency, velocity, position, loudness and pulse rate are initialed..
- Step 2. Exploration: The bats move position based on the behavior of echolocation. All bats explore new location by frequency and velocity according to Eq. (1) and Eq. (2), respectively. In addition, the step of exploration in this study also considers the echo time as depicted in Eq. (3) to adjust the velocity of bats. Hence, the velocity by Eq. (2) is adjusted again by echo time as Eq. (4) to

explore the new location as Eq. (5). If the new location is better than their own position (x_i^{t-1}), the bat will update their own position (x_i^t) with the new found location (l_i^t) according to Eq. (6). If the updated position better than the current best bat, the bat will be the newest the current best bat as shown in Eq. (7).

$$f_i = f_{min} + (f_{max} - f_{min})\beta \tag{1}$$

$$v_i^t = v_i^{t-1} + (x_i^{t-1} - x_*)f_i \tag{2}$$

$$T_i = \frac{2(x_i^{t-1} - x_*)}{v}, V = 340(m/s) \tag{3}$$

$$v_i^t = (v_i^{t-1} \times T_i) + v_i^t \tag{4}$$

$$l_i^t = x_i^{t-1} + v_i^t \tag{5}$$

$$x_i^t = \begin{cases} l_i^t, & \text{if } f(l_i^t) < f(x_i^{t-1}) \\ x_i^{t-1}, & \text{otherwise} \end{cases} \tag{6}$$

$$x_* = x_i^t, \text{ if } f(x_i^t) < f(x_*) \tag{7}$$

where $\beta \in [0,1]$ is a random vector with uniform distribution. The variable, x_* , is the current global best bat (solution) which is located after a comparison of all solutions discovered by n bats. i is the i^{th} bat. T_i is the echo time of bat between it and current global best bat. V is the propagation of sound in temperature is 25°C. t is the current evolutionary iteration. The velocity (v_i^t) is from frequency. Then, the velocity (v_i^t) is obtained by velocity (v_i^{t-1}) and echo time (T_i). l_i^t is the new location of bat i by past position (x_i^{t-1}) and current velocity (v_i^t) in iteration t . x_* is the current global best location.

Step 3. Local search: The main purpose of local search is to attempt to find the better location near current global best bat. A random number (r_p) is used to determine whether a bat begins local search or not. The bat will follow the procedure of local search defined in Eq. (8) if the random number is greater than the rate of pulse emission (r_i) of bat i . Hence, the probability of local search is r_p for each bat. If the new location found by local search is better than their own position (x_i^{t-1}), the bat will update their own position (x_i^t) with the new found location (l_i^t) according to Eq. (6). If the updated position better than the current best bat, the bat will be the newest the current best bat as shown in Eq. (7).

$$l_i^t = x_* + (x_i^t - x_*) \times T_i \tag{8}$$

Step 4. global search: In this stage, in order to avoid the falling into loach optimum solution, the bat can generate a new location that comprises random location and position of current best bat and depends on a threshold (T_λ) as Eq. (9). λ is a random number between 0 to 1.

$$l_{i,j}^t = \begin{cases} \text{random}, & \lambda > T_\lambda \\ x_{*,j}, & \text{otherwise} \end{cases} \tag{9}$$

Step 5. Loudness and pulse emission: The loudness and pulse emission of bats will be updated according to Eq. (10) when (1) the quality of l_i^t generated by flying randomly is better than the quality of the current global best solution (x_*). And (2) the random number (r_A) between 0 and 1 is smaller than the loudness (A_i) of bat i . At the same time, the bat position is also accepted as a new location as follows,

$$A_i^t = \alpha A_i^{t-1}, \quad r_i^t = r_i^0 [1 - \exp(-\gamma t)] \quad (10)$$

$$A_i^t \rightarrow 0, \quad r_i^t \rightarrow r_i^0, \quad \text{as } t \rightarrow \infty \quad (11)$$

where α and γ are constants and assigned to be 0.9 [12, 13] in order to simplify the implemented simulations. Initially, each bat should have different loudness and emission pulse rate. When the iterations increase gradually, according to Eq. (11), these two parameters will slowly approach to zero and initial emission rate (r_i^0), respectively. The loudness and emission pulse rate will be updated only if the new position of the current new bat are improved, which means that the bats are moving towards the optimal solution.

Step 6. Termination: All bats carry out step 2~step 5 to indicate that iteration has been finished. All bats will be back to the step 2 to start next iteration repeatedly until the condition of termination is satisfied. The final global best bat is the optimal solution of this optimization problem.

3 An Echo-Aided Bat Algorithm to Construct Topology in Wireless Sensor Networks

3.1 Scenario Description and Operation Model of Network Simulation

The sensors are grouped into several clusters to build a hierarchical network in WSN. In each cluster, sensors are classified to nodes and cluster head. In this study, the cluster is assumed to form a full-connectivity network topology by logical links as fig. x. Any sensor can directly communicate with other sensors in a cluster and corresponding cluster head. However, direct communication causes the serious quality deterioration and energy consumption due to interference. In order to overcome this problem, spanning tree topology is adopted to manage these logical links in each cluster. In this study, the spanning tree topology is created by the proposed EABA. The elaborative procedures of the created spanning tree topology with EABA are described in Section 3.2. In spanning tree topology, each node has only a parent node and communicates with other sensors through physical link. For an example in fig. x, in fig. 1-b, the node B is the parent node of node A and node C. Hence, the node A and node C must count on node B to deliver the data to sink. On the other hand, the node D has to forward the data from its child node, node E and node F to cluster head. As a whole, the problems of interference and energy consumption will be efficiently improved by spanning tree topology. Finally, the constructed network topology of spanning tree is used to estimate the network lifetime. The detail operations of simulated network are explained in next subsection.

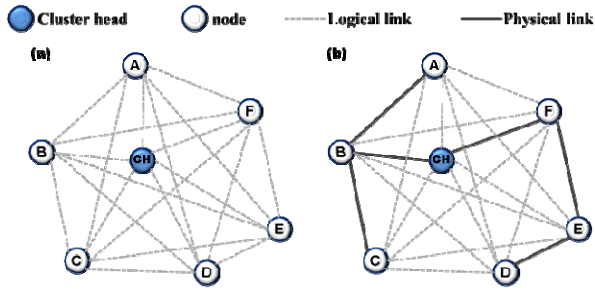


Fig. 1. The diagrams of network topology with (a) full-connectivity and (b) spanning tree in a cluster

The network simulation is built by several procedures: network initialization, Hierarchical network construction, spanning tree route created by EABA, set physical link and data transmission and shown in fig. x. In this subsection, the procedures and operation models of simulated network will be explained in depth. Firstly, in the network initialization, the sensors are deployed by uniform distribution in a region. In the hierarchical network construction, the network is grouped into several clusters by clustering technique to construct the hierarchical network. C is the number of clusters. Each cluster is comprised a cluster head and nodes. Each node has a corresponsive cluster head in allocated cluster. N_k is the number of sensors with nodes and cluster head in cluster k . Then, all sensors set the logical links to form a mesh network topology with full-connectivity. Subsequently, the designed EABA is applied to create the spanning tree network topology with minimum cost for each cluster. In this simulated network, the cost is defined the distance between sensors. For set physical link, in a cluster with spanning tree topology, all nodes only have a parent node to transmit data by physical link to corresponsive cluster head.

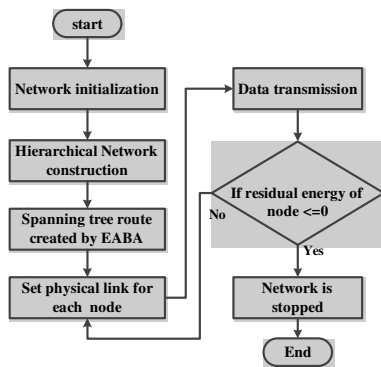


Fig. 2. The flowchart of the simulated network operation model

This network simulation focuses in internal part of a cluster. In data transmission, the data route between nodes and cluster head is established, the nodes respectively

take over different status: end node and relay node. The end node without any child nodes only transmits data to their parent node. The relay node needs forward data from its child nodes and transmits data to its parent node. The cluster head only receive data. Therefore, the needed power consumption is different according to the status of sensors. The power consumption depends on the different status and the needed power consumption is presented as following,

- End node:

$$E_{tx}(k, d) = E_{elec} \cdot l + E_{amp} \cdot l \cdot d^2 = lE_{elec} + lE_{amp}d^2 \quad (12)$$

- Relay node:

$$E_{tx}(k, d) = l(E_{elec} + E_{amp}d^2) + lE_{elec} \quad (13)$$

- Cluster head:

$$E_{rx}(k) = E_{elec} \cdot l \quad (14)$$

where E_{tx} and E_{rx} are the power consumption of l bits data transmission and reception respectively. E_{elec} is the power needed of one bit data transmission or reception in transmitter or receiver circuitry. E_{amp} is the energy needed for transmitter amplifier. d is the distance between two nodes. Moreover, the nodes deliver packet with length l to cluster head in a random time (t) continuously until the energy of node is exhausted. The network operation is stopped to estimate the network lifetime (T) when the energy of any node is exhaustion.

3.2 An Echo-Aide Bat Algorithm to Establish Spanning Tree Topology of Network

An application that construct a network topology with spanning tree by EABA in this study as shown is fig. 3, and this constructed network topology is used to estimate the network lifetime. This operation of this application is divided into four parts: solution presentation, design of fitness function, evolutionary procedures of EABA and a mechanism of activation advancement. In these four parts, evolutionary procedures of EABA was described in Section 2, the procedures of other parts are expressed as below.

- Solution presentation

In this application, each bat presents a solution of spanning tree topology indicated as $= \{e_1, e_2, e_3, \dots, e_\varepsilon\}$. For an example, a cluster can be presented N sensors and an edge set E , i.e., $N_k = \{N, E\}$ in which $N = \{n_1, n_2, n_3, \dots, n_{|N|}\}$ and $E = \{e_{a,b}\}$ where $|E|$ equal to $(N(N-1))/2$ because the degree of node is $N-1$ in full-connectively network. The edge, $e_{a,b}$ stands for the logical link between node n_a and n_b . The numbers of logical links and physical links are ε and ξ respectively, $\varepsilon = |E|$ and $E = \{e_1, e_2, e_3, \dots, e_\varepsilon\}$, $\xi = N - 1$. The length of solution is ξ .

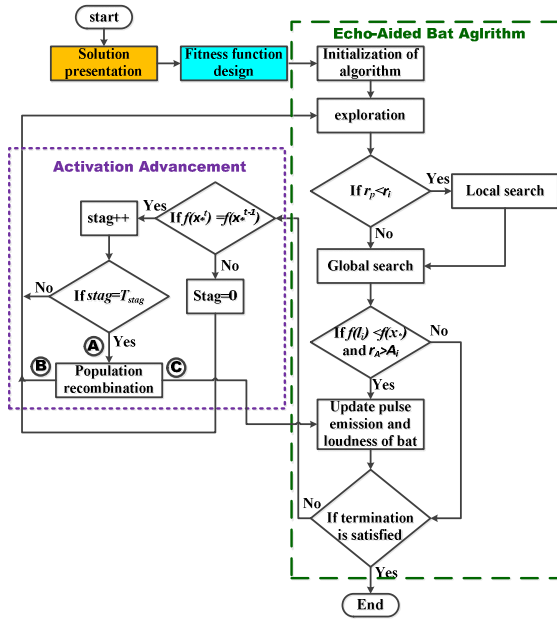


Fig. 3. The flowchart of spanning tree network topology constructed by EABA

➤ Fitness function

Each bat has an own fitness evaluated by the defined fitness function. the fitness function is designed for the spanning tree with minimum cost and used to evaluate the solution quality. Hence, the smaller fitness of bat is, the better quality of bat is. The fitness function for minimum cost is designed as Eq. (15), Eq. (16) and Eq. (17).

$$f(x) = \min \sum_{i=1}^{\xi} d(e_i) \tag{15}$$

$$d(e_i) = \sum_{j=1}^2 \sqrt{n_{a,j} - n_{b,j}}, \quad \text{for } a, b = 1, 2, 3, \dots, N, i=1, 2, 3, \dots, \epsilon \tag{16}$$

$$f(x) = \max_{cost} \sum_{i=1}^{\epsilon} d(e_i) \tag{17}$$

where $f(x)$ is the fitness of bat. e_i is the cost of edge i . $d(e_i)$ is the distance between the node a and node b . There two examination conditions are (1) these edges included in a structure of bat do not cause loop and (2) the edges do not be selected repetitively. If the bats violate the examination conditions, the fitness of bat is calculated by Eq. (17) because the solutions of these kind bats are unable to create a proper spanning tree topology. On the contrary, the bat will obtain a fitness by Eq. (15) if the bats satisfy the examination conditions. These bats will be the candidate of global best solution.

➤ Activation advancement

In this operation, the purpose is to enhance the ability of search for bats. Vigorous search ability benefits the bats to escape the local optimum solutions. And, The global optimum solution will be found quickly. In order to achieve this purpose, a mechanism of population recombination is designed to enhance the search ability of bats shown in fig. 4. When the quality of the global optimum solution does not be improved in a period of iteration, the mechanism of population recombination is triggered. If the fitness of bat is greater than or equal to the T_{qual} depended on the quality ratio (r_q) shown as Eq. (18), the bat should track the current best bat to meliorate its own quality as Eq. (19) and Eq. (20). On the contrary, the bat can learn the better information from other bats to enhance the self-quality according to Eq. (21), when the fitness of bat is less than or equal to the T_{qual} . R is random number with 0 or 1.

$$T_{qual} = f(x_*) + (f(x_*) + r_q) \tag{18}$$

$$v_{i,j}^t = \sigma v_{i,j}^t \quad \sigma \in [-2, 2] \tag{19}$$

$$l_i = x_* + v_i \tag{20}$$

$$l_{i,j} = \begin{cases} l_{i,j}, & \text{if } R = 1 \\ j_{k,j}, & \text{if } R = 0 \end{cases} \tag{21}$$

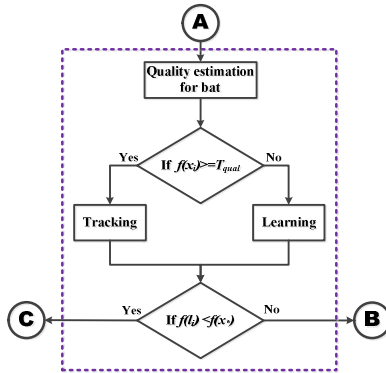


Fig. 4. The flowchart of population recombination in activation advancement

4 Simulation Results

In this study, an echo-aided bat algorithm is applied to construct the network topology of spanning tree. In proposed EABA, there are 40 bats is randomly generated to search the global optimum solution. The performance evaluation of proposed algorithm and the lifetime estimation of constructed network are illustrated as follows.

4.1 Performance Evaluation of the Proposed EABA

In this simulation, there are many scenarios with different node number to evaluate performance of proposed EABA. In this application of spanning tree construction, node number affects the solution space. When the deployed node increases one, the solution space will be increased by ten times. The set node number in this simulation and the solution space as shown in fig. 5. The greater solution space causes the difficulty of find global optimum solution so that the successful rate (SR) is worsened. Each scenario is executed 1000 rounds to derive the statistic and SR. For simulation numeric of SR, in each scenario, the global optimum solution can be found by the proposed EABA. In case of node number is less than 8, the SR almost achieve 100%. However, when the node number more than 8, the SR of find global optimum solution is obvious deterioration due the extensive solution space. The SR is only 41.7% and 36.7 % in the number of nodes is 9 and 10 respectively.

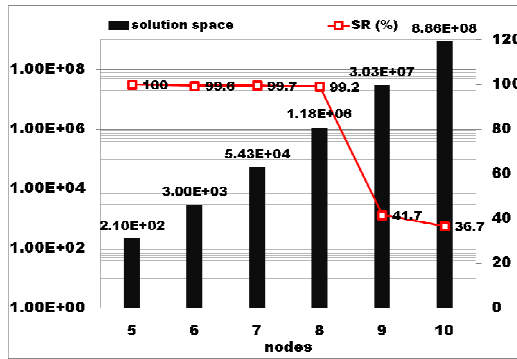


Fig. 5. The solution space and SR in different numbers of node

Although, the SR in the cases of 9 nodes and 10 nodes is not good enough, the performance of algorithm does not disapprove. The statistic of each scenario is shown in table x. the Min. of all scenarios match the global optimum solution to indicate that the proposed EABA has the capability to find the global optimum solution. Besides, in these cases of SR are about 100%, the error of Mean between Min. and standard deviation (STD) are highly slight. This proves that the proposed EABA can discover steadily and accurately the global optimum solution in these scenarios. For the cases of 9 nodes and 10 nodes, the error of Mean between Min. and STD are greater than other cases to deteriorate the SR. Hence, the search capability of proposed EABA is not stabile enough. This situation implies that the bats fall into the local optimum solution in some evolutions. The capability of escape from local optimum solution should be reinforced for bats in the proposed EABA.

Table 1. The statistic of EABA performance in different scenarios

number of node	5	6	7	8	9	10
Mean	300.96 [†]	333.18 [†]	294.11 [†]	364.47 [†]	461.29 [†]	485.02
Median	300.96 [†]	333.11 [†]	293.73 [†]	364.22 [†]	463.32 [†]	479.97
Mode	300.96 [†]	333.11 [†]	293.73 [†]	364.22 [†]	454.49 [†]	473.45
standard deviation	0.00 [†]	1.13 [†]	2.33 [†]	2.96 [†]	7.54 [†]	15.15
Min.	300.96 [†]	333.11 [†]	293.73 [†]	364.22 [†]	454.49 [†]	473.45
Max.	300.96 [†]	351.04 [†]	315.90 [†]	411.49 [†]	502.00 [†]	558.80

4.2 Lifetime Estimation of the Constructed Network Topology

The proposed EABA presents satisfied performance in a great deal of scenarios designed in this study. A network topology of spanning tree is constructed by the found global optimum solution in the case of 10 nodes. This network topology is used to estimate the network lifetime in this simulation. In this simulation, the initial energy is 1000 J for all sensors. A sensor excluding cluster head is randomly selected to transmit one packet or two packets to cluster head through the data transmission route in each times of data transmission. The length of packet is 512 bits. And, the transmission power, reception power and amplifier power are 0.0104 J, 0.086 J and 0.25 J respectively. The power consumption is calculated by Eq. (12), Eq. (13) and Eq. (14). Once the residual energy of one node is equal to 0, the network stop transmission to obtain the residual energy of other nodes and the amount of received data in cluster

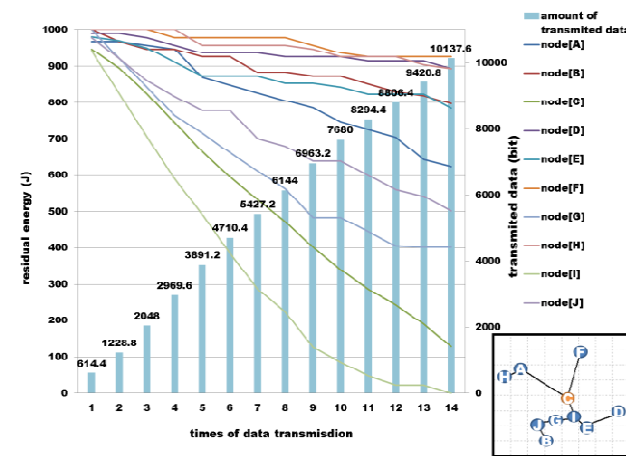


Fig. 6. Network lifetime simulation of constructed network topology by EABA

head. In this network topology, the sensor C is selected as cluster head to decide the status of other sensors following, sensor A, I, G, E, J are the relay nodes. The end nodes are sensor H, B, F and D. This simulated network topology of spanning tree is executed 5 rounds of transmission to estimate the network lifetime as shown in fig. 6. The energy of sensor I is exhausted in fourteenth time of data transmission because it is the relay nodes and forward the data for 5 sensors. Then, the cluster head received data continuously so that its power consumption is more than other sensors excluding sensor I. Hence, there are total about 10137 bits to be deliver to cluster head (sensor C) within 14 times of data transmission.

5 Conclusions

The framework includes (1) EABA is applied to construct the network topology of spanning tree, and (2) this constructed network topology is used to estimate the network lifetime is created in this study. In the construction of network topology with spanning tree, EABA presents satisfied performance in a large number of scenarios designed in this simulation. For all scenarios, the global optimum solution can be found by the proposed EABA, but the SR of find global optimum solution is not very perfect in few scenarios. Hence, the capability of global search is still strengthened to avoid falling into local optimum solution for all bats in discrete problems. Afterward, the effect of different between continuous and discrete problems for evolutionary computing is discussed in depth in future. Besides, the network topology of spanning tree constructed by global optimum solution is used to estimate the network lifetime thought simulation of network transmission. This created framework supports the performance evaluation before WSN deployment. However, in addition to the network topology, unbalanced workload is another important factor to influence the network performance. The issue of unbalanced workload is considered to increase the amount of data transmission in future.

Acknowledgement. The authors would like to express their sincere thanks to the National Science Council, Taiwan (ROC), for financial support under the grants NSC 102-2221-E-151 -039 - , NSC 102-2622-E-151 -004 -CC3 and NSC 102-2218-E-151 -005 -.

References

1. Brazil, M.N., Ras, C.J., Thomas, D.A.: Relay augmentation for lifetime extension of wireless sensor networks. *IET Wireless Sensors System* 3(2), 145–152 (2013)
2. Incel, O.D., Ghosh, A., Krishnamachari, B., Chintalapudi, K.: Fast Data Collection in Tree-Based Wireless Sensor Networks. *IEEE Transactions on Mobile Computing* 11(1), 86–99 (2012)
3. Paschalidis, I.C., Li, B.B.: Energy Optimized Topologies for Distributed Averaging in Wireless Sensor Networks. *IEEE Transactions on Automatic Control* 56(10), 2290–2304 (2011)

4. Bui, T.N., Deng, X.H., Zrncic, C.M.: An Improved Ant-Based Algorithm for the Degree-Constrained Minimum Spanning Tree Problem. *IEEE Transactions on Evolutionary Computation* 16(2), 266–278 (2012)
5. Ernst, A.T.: A hybrid Lagrangian Particle Swarm Optimization Algorithm for the degree-constrained minimum spanning tree problem. In: *Proceeding of IEEE Congress on Evolutionary Computation (CEC 2010)*, pp.1–8 (2010)
6. Chen, Y.T., Lee, T.F., Horng, M.F., Pan, J.S.: An Echo-Aided Bat Algorithm to Support Measurable Movement for Optimization Efficiency. In: *Proceeding of IEEE International Conference on Systems, Man, and Cybernetics (SMC 2013)*, pp. 806–811 (2013)
7. Yang, X.-S.: A New Metaheuristic Bat-Inspired Algorithm. In: González, J.R., Pelta, D.A., Cruz, C., Terrazas, G., Krasnogor, N. (eds.) *NICSO 2010. SCI*, vol. 284, pp. 65–74. Springer, Heidelberg (2010)
8. Yang, X.S.: Bat Algorithm for multi-objective optimization. *International Journal of Bio-Inspired Computation* 3(5), 267–274 (2011)

Part VIII
**Cross-Discipline Techniques in Signal
Processing and Networking**

Design of Triple-Band Planar Dipole Antenna

Yuh-Yih Lu^{1,*}, Jun-Yi Guo¹, Kai-Lun Chung¹, and Hsiang-Cheh Huang²

¹Department of Electrical Engineering,
Minghsin University of Science and Technology, Hsinchu 304, Taiwan, R.O.C.
yylu@must.edu.tw

²Department of Electrical Engineering,
National University of Kaohsiung, Kaohsiung 811, Taiwan, R.O.C.
hch.nuk@gmail.com

Abstract. A new low-profile planar symmetric dipole antenna is proposed for triple-band wireless communication application. Three symmetric arms are etched on the metallic layer of a single sided printed circuit board to form the planar dipole antenna. The dimensions of symmetric arms are changed to design and fabricate the antenna which can be operated at 2.45/3.5/5.8 GHz successfully. We use IE3D software to design this triple-band planar antenna and choose the better parameters to manufacture the proposed antenna. The influences of dimension parameter of the proposed antenna on the resonant frequency and impedance bandwidth are described. The proposed antenna with the small volume of 40.6mm×13mm×0.6mm has been fabricated and this antenna can be used in WLAN frequency band.

Keywords: symmetric arm, planar dipole antenna, WLAN.

1 Introduction

It is well known that compact size, lower cost and easy fabrication are important factors to design antenna that can be used in wireless communication. Planar antennas possess the attractive features. Hence, many studies about planar antennas had been proposed and widely used in Wireless Local Area Network (WLAN), Radio Frequency Identification (RFID), Worldwide Interoperability for Microwave Access (WiMAX) and Ultra Wide Band (UWB) systems [1-9]. In modern wireless communication devices, these information devices should be capable of operating at multiple frequency bands. Therefore, researches have been reported for multi-band planar antenna [10-12].

In this study, a simple uniplane dipole antenna with symmetric arms is proposed. The return loss, resonant frequency, impedance bandwidth, and radiation pattern are obtained from IE3D simulations. The lower, middle and upper arms of the proposed antenna control the resonant frequency of the triple-band dipole antenna. The planar dipole antennas designed with the symmetric arms excite the resonant frequency that can be used for WLAN applications. The suitable geometric parameters of the symmetric arms are chosen to fabricate the proposed antenna which can operate at

* Corresponding author.

2.45/3.5/5.8 GHz. Therefore, a simple planar symmetric dipole antenna with small size of 40.6mm×13mm×0.6mm is presented in this paper. The proposed antenna can be built on a single sided printed circuit board. The single metal layer structure is suitable for mass production and reduces the manufacturing cost.

2 Antenna Design

The proposed antenna has a compact size of 40.6mm×13mm. The proposed planar dipole antenna structure is printed on a single metallic layer of FR4 dielectric substrate which has permittivity of 4.4 and thickness of 0.6mm. The configuration of this proposed antenna is depicted in Fig.1. In this figure, three symmetric arms are etched on the metallic layer to create the operating frequency bands. Points A and B are the feeding points of the planar dipole antenna. We adjust the lower arm length parameter $L1$, middle arm length parameter $L2$ and upper arm corner parameter S to observe the variations with respect to the resonant frequency and impedance bandwidth of the proposed antennas. The dimension parameters of the proposed antenna shown in Fig.1 are listed below: $W=10\text{mm}$, $L3=14.5\text{mm}$, $L4=3\text{mm}$, $L5=2\text{mm}$, $G=1\text{mm}$, $W1=2\text{mm}$, $W2=2\text{mm}$, $W3=2\text{mm}$, $W4=2\text{mm}$, $W5=3\text{mm}$, $W6=7\text{mm}$, $W7=1\text{mm}$. The 50 ohm coaxial connector was adopted for testing.

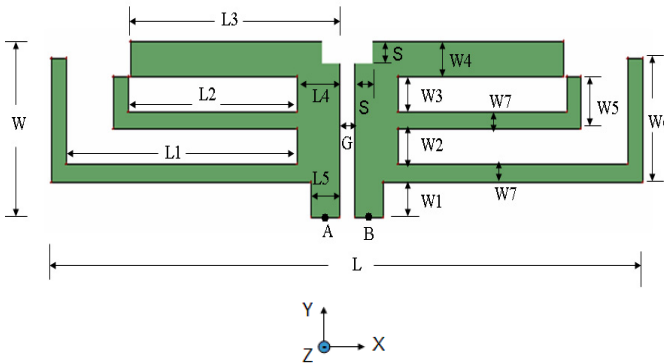


Fig. 1. Geometry of the proposed planar dipole antenna

3 The Simulations

We adopted various dimension parameters $L1$, $L2$ and S shown in Fig.1 of the planar symmetric dipole antenna to observe the characteristics of the proposed antenna. The numerical simulation and analysis for the proposed antennas are performed using IE3D simulation software. The simulated curves of return loss against frequency for varying the lower arm parameter $L1$ of the proposed antenna with $L2=11.7\text{mm}$ and $S=1\text{mm}$ are shown in Fig.2. From this figure, three obvious operating frequency bands are observed and the lower resonant frequency is shifted to lower frequency

with increasing the value of L_1 . The simulated curves of return loss against frequency for varying the middle arm parameter L_2 of the proposed antenna with $L_1=16\text{mm}$ and $S=1\text{mm}$ and varying the upper arm parameter S of the proposed antenna with $L_1=16\text{mm}$ and $L_2=11.7\text{mm}$ are shown in Fig.3 and Fig.4, respectively.

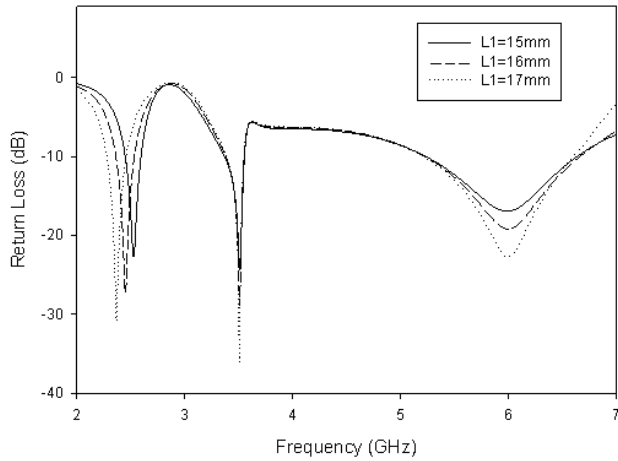


Fig. 2. Simulated curves of return loss against frequency for varying L_1 of the proposed antenna

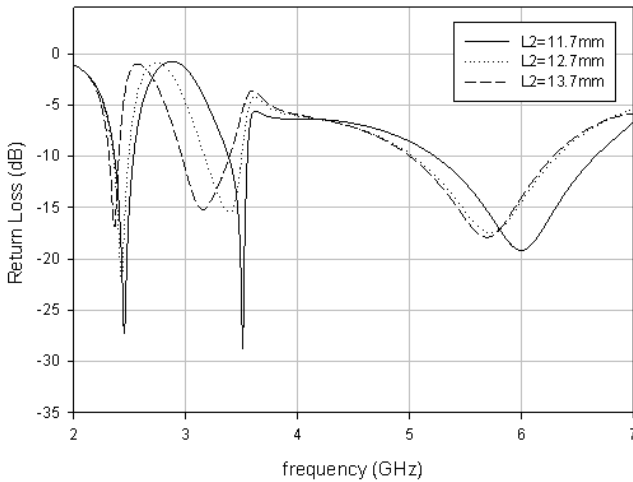


Fig. 3. Simulated curves of return loss against frequency for varying L_2 of the proposed antenna

From Fig. 3 and Fig. 4, three obvious operating frequency bands are also observed. The lower, middle and upper resonant frequencies are shifted to lower frequency with increasing the value of L_2 shown in figure 3. As shown in Fig.4, the lower and middle resonant frequencies are nearly unchanged while the upper resonant frequency is shifted to higher frequency with increasing the value of S . The lower, middle and upper operating frequency band simulated results are listed in Table 1. These simulated results include resonant frequency (f_c), return loss (RL), and impedance bandwidth (BW).

In order to design the proposed antenna that can be used in 2.45/3.5/5.8 GHz, we choose $L_1=16\text{mm}$, $L_2=11.7\text{mm}$ and $S=1\text{mm}$ to fabricate the triple-band planar symmetric dipole antenna. The radiation patterns are computed using IE3D software for the proposed antenna with $L_1=16\text{mm}$, $L_2=11.7\text{mm}$ and $S=1\text{mm}$. Fig.5 shows the simulated radiation patterns of this antenna. The computed peak gains at the operating frequency obtained from the radiation patterns are classified as Table 2. It can be seen that the radiation patterns are almost omnidirectional in the y - z plane as shown in Fig.5. From the simulation results, it is easy to find that the simulated return loss, impedance bandwidth and peak gain at the lower, middle and upper frequency bands show good performance and can be used at 2.45/3.5/5.8 GHz.

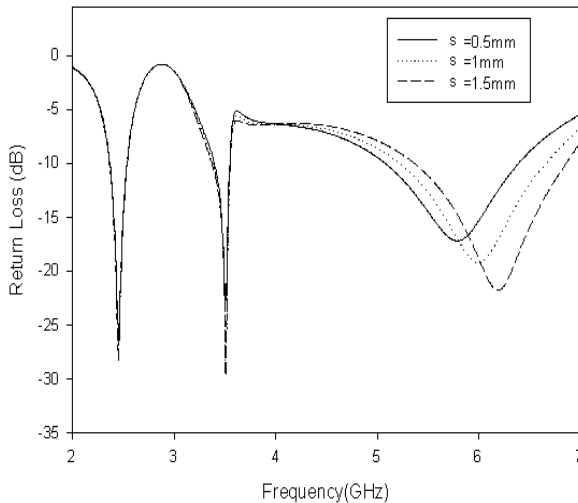


Fig. 4. Simulated curves of return loss against frequency for varying S of the proposed antenna

4 Experimental Results and Discussion

From the simulation results, we use the appropriate geometric parameters to fabricate the proposed antenna. To reach the operating frequencies covering 2.45/3.5/5.8 GHz, we choose $L_1=16\text{mm}$, $L_2=11.7\text{mm}$ and $S=1\text{mm}$ to fabricate the desired antenna. The photography of fabricated antenna is shown in Fig.6. The curves of return loss against

frequency of the simulated and fabricated antenna are illustrated in Fig.7. The simulated and measured results are listed in Table 3. From these data, we observe that the trend of simulated and measured operating frequency band and return loss are in good agreement. The measured impedance bandwidths of the fabricated antenna for return loss less than -10dB at lower, middle and upper frequency band are 320MHz, 230MHz and 3160MHz, respectively. The measured return loss and impedance bandwidth of the fabricated antenna show better performance than that in simulation condition.

Table 1. Simulated results for varying dimension parameters of the proposed antenna

L1 (mm)	L2 (mm)	S (mm)	Frequency band	f_c (GHz)	RL (dB)	BW (MHz)
15	11.7	1	lower	2.53	-22.6	145
			middle	3.5	-24.2	184
			upper	6	-16.8	1398
16	11.7	1	lower	2.45	-27	170
			middle	3.5	-28.6	170
			upper	6	-19.1	1435
17	11.7	1	lower	2.36	-30.8	189
			middle	3.5	-35.9	161
			upper	5.96	-22.6	1389
16	12.7	1	lower	2.43	-21.7	138.5
			middle	3.38	-15.4	330
			upper	5.71	-17.5	1285
16	13.7	1	lower	2.37	-16.6	88
			middle	3.2	-15.1	427
			upper	5.65	-17.8	1282
16	11.7	0.5	lower	2.45	-26.3	169
			middle	3.5	-23.2	149
			upper	5.78	-17.1	1308
16	11.7	1.5	lower	2.44	-28.1	171
			middle	3.5	-29.5	179
			upper	6.2	-21.7	1491

The measured radiation patterns of the fabricated antenna at 2.45/3.5/5.8 GHz are shown in Fig.8. The measured peak gains for testing frequencies at x-z and y-z plane of the fabricated antenna are listed in Table 4. There are discrepancies between the computed and measured results which may occur because of the effect of the coaxial connector soldering process and fabrication tolerance.

From Fig.8, it can be observed that the radiation patterns are almost omnidirectional in the y-z plane. The omnidirectional antenna radiation pattern indicates that the fabricated antenna is good for mobile devices.

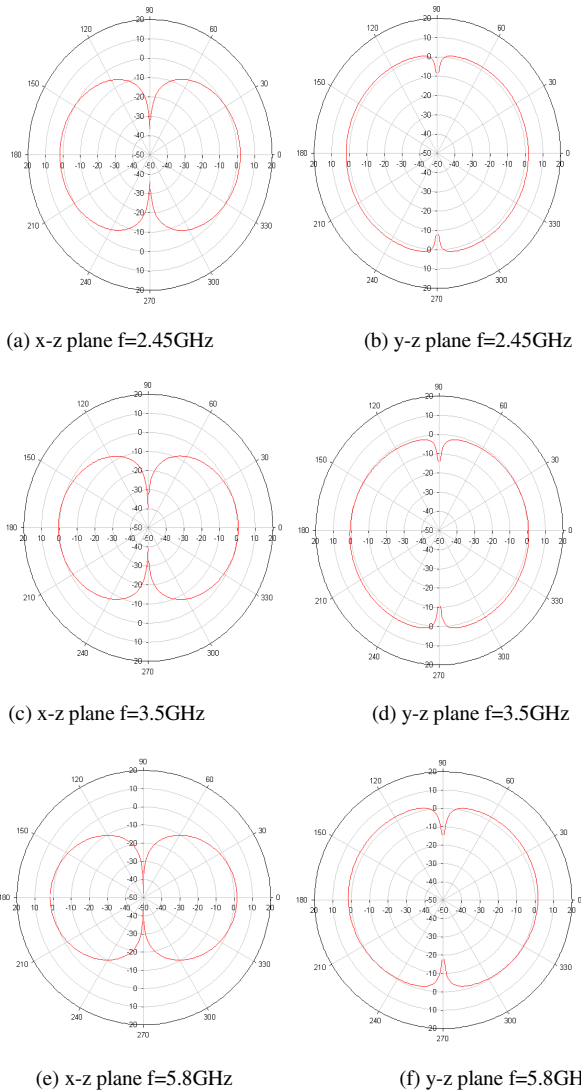


Fig. 5. Simulated radiation patterns of the proposed antenna

Table 2. Simulated results of the proposed antenna at operating frequency

f (GHz)	x-z plane GAIN (dBi)	y-z plane GAIN (dBi)
2.45	1.79	1.97
3.5	0.64	2.08
5.8	1.57	2.35

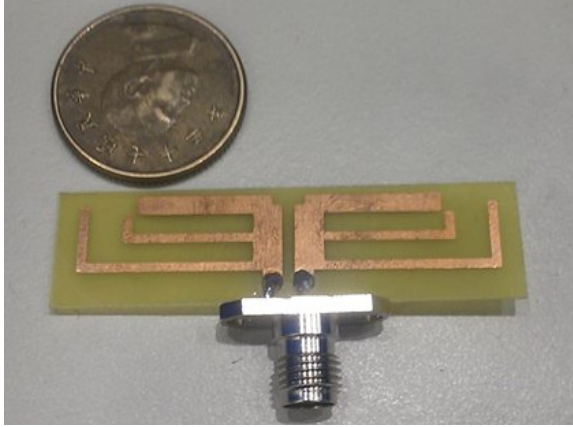


Fig. 6. Photography of fabricated planar dipole antenna

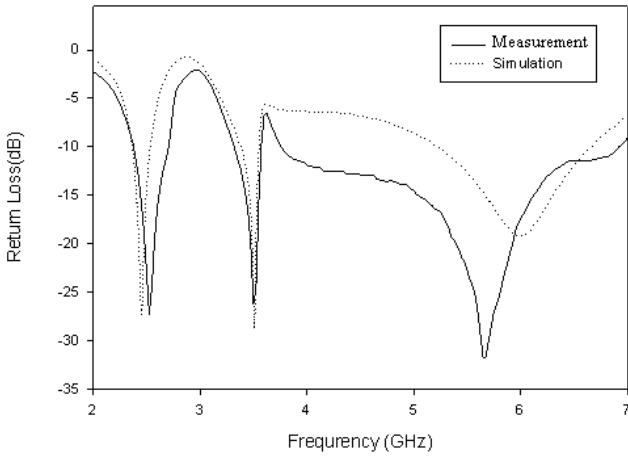
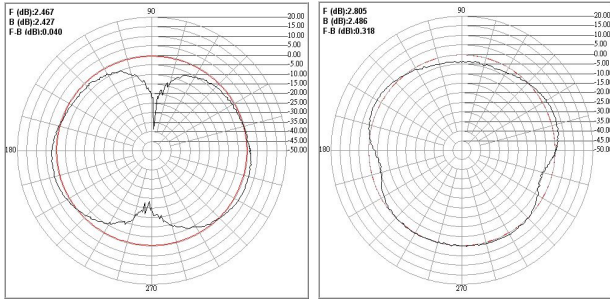


Fig. 7. Simulated and measured return loss of the proposed antenna

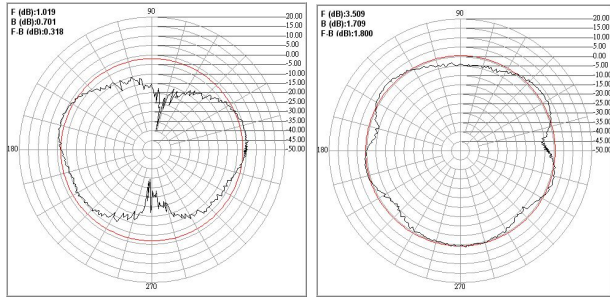
Table 3. Simulated and measured results of the proposed antenna

Condition	f_c (GHz)	RL (dB)	BW (MHz)
Simulation	2.45	-27	170
	3.5	-28.6	170
	6	-19.1	1435
Measurement	2.52	-27.3	320
	3.5	-26	230
	5.65	-31.6	3160



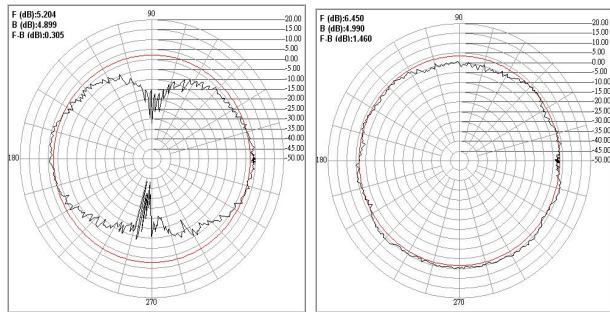
(a) x-z plane $f=2.45\text{GHz}$

(b) y-z plane $f=2.45\text{GHz}$



(c) x-z plane $f=3.5\text{GHz}$

(d) y-z plane $f=3.5\text{GHz}$



(e) x-z plane $f=5.8\text{GHz}$

(f) y-z plane $f=5.8\text{GHz}$

Fig. 8. Measured radiation patterns of the proposed antenna

Table 4. Measured results of the fabricated antenna at operating frequency

f (GHz)	x-z plane GAIN (dBi)	y-z plane GAIN (dBi)
2.45	2.46	2.8
3.5	1.02	3.5
5.8	5.2	6.45

5 Conclusions

In this study, the fabricated triple-band planar dipole antenna exhibits simple structure and small size on a 40.6mm×13mm substrate. It can excite the frequency band that can be used in 2.45/3.5/5.8 GHz. Therefore, carefully choose the designed parameter of the symmetric arm would implement the suitable antenna and perform another type of planar dipole antenna. The fabricated antenna can be built as an on-board antenna. This will reduce the manufacturing cost of a device.

References

1. Khaleghi, A.: Dualband Meander Line Antenna for Wireless LAN Communication. *IEEE Trans. Antennas and Propagation* 55, 1004–1009 (2007)
2. Lu, Y.Y., Wei, S.C., Huang, H.C.: Design of RFID Antenna for 2.45GHz Applications. In: *Proceedings of ICICIC*, pp. 601–604 (2009)
3. Chen, Y.C., Chiu, P.T., Tsai, J.M., Chen, M.D.: A Compact Triple-band Planar Monopole Antenna for WLAN and WiMAX Applications. In: *2013 International Workshop on Antenna Technology (iWAT)*, pp. 311–314 (2013)
4. Dubrovka, F.F., Vasylenko, D.O.: A Bell-Shaped Planar Dipole Antenna. *Ultrawideband and Ultrashort Impulse Signals*, 82–84 (2006)
5. Ayop, O., Rahim, M.K.A., Masri, T.: Planar Dipole Antenna with and without Circular Parasitic Element. In: *Asia-Pacific Conference on Applied Electromagnetics*, pp. 1–4 (2007)
6. Zhang, J.P., Xu, Y.S., Wang, W.D.: Ultra-wideband Microstrip-fed Planar Elliptical Dipole Antenna. *Electronics Letters* 42, 144–145 (2006)
7. Lee, H.R., Woo, J.M.: Asymmetric Planar Dipole Antenna on the Surface of Conducting Plane for RFID Tag. In: *Asia Pacific Microwave Conference*, pp. 633–636 (2009)
8. Gupta, S., Ramesh, M., Kalghatgi, A.T.: Design of Optimized CPW Fed Monopole Antenna for UWB Applications. In: *Proceedings of Asia-Pacific Microwave Conference*, p. 12 (2005)
9. Chair, R., Kishk, A.A., Lee, K.F.: Ultrawide-band Coplanar Waveguide-fed Rectangular Slot Antenna. *IEEE Antennas and Wireless Propagation Letters* 3, 227–229 (2004)
10. Xu, P., Yan, Z.H., Wang, C.: Multi-band Modified Fork-shaped Monopole Antenna with Dual L-shaped Parasitic Plane. *Electronics Letters* 47, 364–365 (2011)
11. Wu, C.M., Chiu, C.N., Hsu, C.K.: A New Nonuniform Meandered and Fork-Type Grounded Antenna for Triple-Band WLAN Applications. *IEEE Antennas and Wireless Propagation Letters* 5, 346–348 (2006)
12. Wong, K.L.: *Planar Antennas for Wireless Communications*. Wiley, Hoboken (2003)

Solar Irradiance Estimation Using the Echo State Network and the Flexible Neural Tree

Sebastián Basterrech and Tomáš Buriánek

IT4Innovations

VŠB–Technical University of Ostrava,

Czech Republic

{Sebastian.Basterrech.Tiscordio,Tomas.Burianek.St1}@vsb.cz

Abstract. Two popular models for solving temporal learning problems are the *Flexible Neural Tree (FNT)* and the *Echo State Network (ESN)*. Both models belong to the the Neural Network area. The ESN is based in the projection of a recurrent neural network to model the temporal dependencies of the data. The FNT uses heuristic techniques for finding a tree topology and its parameters. There are several examples in the Machine Learning literature that shown the success for solving learning tasks of both techniques. In this paper, we have studied the performance of these methods in a specific data set about renewable energy.

Keywords: Echo State Network, Flexible Neural Tree, Time-series forecasting, Reservoir Computing, Renewable energy.

1 Introduction

At the beginning of the 2000s, two models for solving temporal learning tasks were presented in the Machine Learning community: *Flexible Neural Tree (FNT)* [1, 2] and the *Echo State Network (ESN)* [3]. The first one is a Feed-Forward Neural Network (NN) with a specific topology defined by optimization using heuristic techniques. The second one, is a three layer recurrent NN where the input layer is full connected with the hidden layer and this one is full connected with the output layer. The circuits are presented only in the hidden layer. The distinguish property of this model is that the hidden layer is fixed during the learning process, only the output weights are updated in the training. Both kind of Neural Networks have been succesfully used in time series modeling and forecasting problems. A forecast problem consists in extracting all possible information about the future considering the information already presented about the past. Most formal, we are interested in some quantity $\mathbf{y}_{\text{target}}(t) \in \mathbb{R}^{N_y}$ where $t = 1, 2, \dots$ (we assume time discrete). Assume that we have observed $\mathbf{y}_{\text{target}}(1), \mathbf{y}_{\text{target}}(2), \mathbf{y}_{\text{target}}(3), \dots, \mathbf{y}_{\text{target}}(n)$, the problem consists in estimating the $\mathbf{y}_{\text{target}}(n+k)$ for some $k > 0$.

In this paper we study these two models as learning predictors on a data set about the energy sources. Specifically, we are interesting in predict the solar irradiance at the current moment using the information of the solar irradiance

of a past sliding windows. The interest of this tasks arises from the fact that the global solar irradiance is an important meteorological variable in the production of renewable energy sources.

The remainder of this paper is organized as follows. We present a description of the Flexible Neural Tree model in the next section. Section 3 contains a brief presentation about Reservoir Computing and the Echo State Network model. We then present our experimental results for this particular temporal task. Finally, the last part presents the article conclusion.

2 The Flexible Neural Tree Model Description

Ten years ago a kind of multilayer feed-forward Neural Network (NN) was introduced in the Machine Learning literature under the name of *Flexible Neural Tree (FNT)* [1, 2]. We can see a FNT as a feed-forward NN with an irregular topology and a specific parametric activation function. The FNT model uses a simple random search method for optimizing the tree parameters, which are the weight connections and the activation function parameters. In order to define the network topology an evolutionary technique is employed. The interest in the FNT stems from the fact that the pattern of connectivity among the neurons is designed following an automatic process. That includes also the number of input variables in the tree. An heuristic procedure is used for finding the best connectivity among the units in the tree. For this purpose have been used the bio-inspired techniques such that: the Probabilistic Incremental Program Evolution (PIPE) [1], *Genetic Programing (GP)* [4–7] and *Ant Programming (AP)* [8]. A second procedure is employed for tuning the neural tree parameters, in this case techniques that *Particle Swarm Optimization (PSO)* [9, 10] and *Differential Evolution (DE)* [11, 12] have been used.

2.1 The Flexible Neuron Instructor

The tree structure is created using a pre-defined instruction set. We follow the notation proposed in [2]. Let $\mathcal{F} = \{+_2, +_3, \dots, +_{N_f}\}$ be a *function set*. Each element of \mathcal{F} denotes a flexible node operator, $+_i$ is an internal vertex instruction with i inputs. The activation function of any unit i is a parametric function with the following form:

$$f(a_i, b_i, x) = e^{-\left(\frac{x-a_i}{b_i}\right)^2}, \quad (1)$$

where a_i and b_i are two adjustable parameters. We denote by \mathcal{T} a set of N_t elements called *terminal set*: $\mathcal{T} = \{x_1, x_2, \dots, x_{N_t}\}$. Let \mathcal{S} be the *instruction set* defined as follows:

$$\mathcal{S} = \mathcal{F} \cup \mathcal{T} = \{+_2, +_3, \dots, +_{N_f}\} \cup \{x_1, x_2, \dots, x_{N_t}\}. \quad (2)$$

For each node $+_i$, there are i random nodes as its inputs, which can be internal (non-leaf) and terminal (leaf) nodes. Let the terminal nodes $\{x_1, \dots, x_i\}$ be inputs of the functional node $+_i$, the total input charge of $+_i$ is defined as:

$$net_i = \sum_{j=1}^i w_j x_j, \quad (3)$$

where w_j is the weight connection between x_j and $+_i$. The output of the node $+_i$ is obtained using the expression (1)

$$out_i = f(a_i, b_i, net_i). \quad (4)$$

When a functional node has as its inputs other functional nodes, we use the expressions (3) and (4) following some strategy for traversing the tree. For instance: a depth-first strategy for traversing the tree topology can be used. Let N be the number of nodes in the tree, as a consequence the entire tree is traversed in $O(N)$ algorithmic time. In order to evaluate the accuracy of the FNT model the *Mean Square Error (MSE)* is considered as the *fitness function* of the heuristic techniques. Another important parameter of the model performance is the number of nodes in the tree. In the case that two trees have equal accuracy level, the smallest is more performant in time.

2.2 The Parameter and Topology Optimization Using Heuristic Techniques

Several strategies have been used to find the optimal neural tree and the optimal parameters embedded in the structure [1, 2, 4, 5, 13]. The model parameters embedded in the tree are the activation function parameters (a_i and b_i , for all $+_i$) and the weight connections between nodes. In this paper, we use *Genetic Programming (GP)* for finding an optimal connectivity among the neurons.

A GP algorithm starts defining an initial population of specific devices. The procedure is iterative, at each epoch it transforms a selection of individuals generating a new generation. This transformation consists in applying some bio-inspired rules to the individuals. A set of individuals are probabilistically selected and a set of genetic rules is applied. In our problem, the individuals of the population are the flexible trees. The operating rules arise from some biological genetic operations, which basically consist of: *reproduction*, *crossover* and *mutation*. The reproduction is the identity operation, a individual i of a generation at time t is also presented at the generation at time $t + 1$. Given two individuals i and j at time t , the crossover consists in generating an individual s at time $t + 1$ using information from i and j . The mutation consists in selecting a tree and realizing one of the following operations: to change a leaf node to another leaf node, to replace a leaf node for a sub-tree, and to replace a functional node by a leaf node.

Each functional node has the parameters a_i and b_i on the activation function given by 1, another kind of adjustable parameters in the system are the weight

connection among the nodes. In this paper, we use *Particle Swarm Optimization (PSO)* [9] for finding these parameters. The PSO algorithm is an evolutionary computation technique based on social behaviors in a simplified social environment. A *swarm* is a set of particles where each particle is characterized by its position and its velocity in a multidimensional space. We denote the position of a particle i with the column N_x -vector $\mathbf{x}^{(i)} = (x_1^{(i)}, \dots, x_{N_x}^{(i)})$ and the velocity of i is defined by the column N_x -vector $\mathbf{v}^{(i)} = (v_1^{(i)}, \dots, v_{N_x}^{(i)})$. Besides, we use auxiliary vectors $\mathbf{p}^{(i)}, \forall i$ and \mathbf{p}^* , each one with dimension $N_x \times 1$. The vector $\mathbf{p}^{(i)}$ denotes the best position of i presented until current iteration. The best swarm position is represented by the vector \mathbf{p}^* . Besides, we use two auxiliary random weights $\mathbf{r}_1^{(i)}$ and $\mathbf{r}_2^{(i)}$ of dimensions $N_x \times 1$, which are randomly initialized in $[0, 1]$ for each particle i . At any iteration t , the simulation of the dynamics among the particles is given by the following expressions [14]:

$$v_j^{(i)}(t+1) = c_0 v_j^{(i)}(t) + c_1 r_{1j}^{(i)}(t) (p_j^{(i)}(t) - x_j^{(i)}(t)) + c_2 r_{2j}^{(i)}(t) (p_j^*(t) - x_j^{(i)}(t)), j \in [1, N_x] \quad (5)$$

and

$$x_j^{(i)}(t+1) = x_j^{(i)}(t) + v_j^{(i)}(t+1), j \in [1, N_x], \quad (6)$$

where the constant c_0 is called the *inertia weight* the constants c_1 and c_2 regulate local and global position of the swarm, respectively. In order to use PSO for estimating the embedded tree parameters, the position $\mathbf{p}^{(i)}$ of the particle i is associated with the embedded parameters in one flexible tree (the weights and $a_j, b_j, \forall j \in \mathcal{F}$). The PSO algorithm return the global optimal position according the fitness function. The relationship between the tree parameters and the particle position is given by:

$$(p_1^{(i)}, \dots, p_{N_x}^{(i)}) = (a_1, \dots, a_{N_f}, b_1, \dots, b_{N_f}, \mathbf{w}), \quad (7)$$

where \mathbf{w} is a vector with the tree weights.

3 Reservoir Computing

At the beginning of 2000s were independently introduced the *Echo State Network (ESN)* [3] and the *Liquid State Machine (LSM)* [15]. The ESN model was developed in the field of Control Systems and Machine Learning and the LSM model arises from the Neurocomputing area. Although, they have different origins both methods share the same basic principle, they have two well-distinguished structures. One part is used for encode temporal information and the another one is employed for supervised training adaptation. The structure for temporal coding the input information is often a Recurrent Neural Network (RNN), which is called *reservoir* in the ESN context. Since 2007 the approach of these two models collide forming a new paradigm referred by the name of *Reservoir Computing (RC)*. RC models are based in the principle that a dynamical system (a fixed

RNN) provides a complex nonlinear transformation of the temporal input patterns, which is sufficient to encode temporal input information and to improve the linear separability of the input data. Therefore, the readout structure is used to extract the features in the reservoir in order to estimate the desired outputs. In practical applications, it is enough to define the readout structure as a simple learning tool, for instance linear regression [16]. Only the readout structure is updated using the training data in the learning process.

Several variations of RC models have been proposed in the literature, such as: Intrinsic Plasticity [17], Backpropagation–decorrelation [18], Decoupled ESN [19], ESN with leaky integrators [20], Evolino [21], a RC method which uses ideas from the Queueing Theory called *Echo State Queueing Network (ESQN)* [22]. RC models have been proven to be a very good alternative of Turing Machines and RNN to model cognitive processing in the neural system as well as to solve time-series problems and forecasting [16]. Some of special application areas of RC methods are Pattern Classification [3], Speech Recognition [23, 24], Intrinsic Computation [17] and Time-series prediction [3, 16, 18, 25–27].

The reason for choosing the ESN model is that it has been widely used in the supervised learning literature. For instance, the best known learning performance on the Mackey–Glass times series benchmark problem was obtained using the ESN model as predictor [21].

3.1 The Echo State Network Model

The ESN model is composed by a reservoir generated with a RNN and the readout structure is a linear regression. In the original model [3], the activation function of the units is the $\tanh(\cdot)$ function and the pattern of connectivity among the neurons is not dense and random. According to empirical experiences is enough to use between 10% and 30% of connections among the units [28]. The number of reservoir units is much larger than the dimensionality of the input space. The reservoir should verify some algebraic conditions, in order to guaranteed the stability of the RNN. These conditions are specified in the *Echo State Property* [3]. This property states that the dynamical system converges given the same inputs, regardless the previous states and past inputs. In practice this property is guaranteed when the weight matrix of reservoir connectivity is appropriately scaled [16]. In this work, we use the batch training ridge linear regression, in order to compute the linear regression parameters [16].

We follow the previous notation concerning the training set. We have an input space \mathbb{R}^{N_x} and an output space \mathbb{R}^{N_y} . The training set is collected in the pairs $(\mathbf{x}(t), \mathbf{y}_{\text{target}}(t))$, $t = 1, \dots, T$. The reservoir information is represented in a N_s -vector (t) ($N_x \ll N_s$), computed as follows:

$$s_m(t) = \tanh\left(w_{m0}^{\text{in}} + \sum_{i=1}^{N_x} w_{mi}^{\text{in}} x_i(t) + \sum_{i=1}^{N_s} w_{mi}^r s_i(t-1)\right), \forall m \in [1, N_s], \quad (8)$$

where the weight connections between input and reservoir nodes are given by a $N_s \times N_x$ weight matrix \mathbf{w}^{in} , the connections among the reservoir neurons are

represented by a $N_s \times N_s$ weight matrix \mathbf{w}^r and a $N_y \times (N_x + N_s)$ weight matrix \mathbf{w}^{out} represents the connections between reservoir and output units. For sake of notation simplicity, we omit the bias term included in the linear regressions. The network outputs (*readouts*) are generated by a linear regression. Thus, the network output y_m is computed as follows:

$$y_m(t) = w_{m0}^{\text{out}} + \sum_{i=1}^{N_x} w_{mi}^{\text{out}} x_i(t) + \sum_{i=1}^{N_s} w_{mi}^{\text{out}} s_i(t), \quad \forall m \in [1, N_y]. \quad (9)$$

4 Experimental Setup and Results

In this study we use real data about solar energy. The data was measured at a power plant located near to Starojicka Lhota, Czech Republic. The data was collected during July 1, 2010 till March 31, 2011. The solar power is an important variable in the renewable energy production. For more details about this data set, see [29]. Below, a preliminary study about FNT for estimating the solar power using this data set was analyzed in [30]. The data set was normalized in $[0, 1]$. The goal is to predict the solar irradiance using information about the history of this variable. The data set contains 36 attributes and 1 response variable. The input features corresponds a sliding window, which explores the seasonal traits of the data set. The input pattern were measured each 10 minutes, the training set comprises 27278 samples (66% of the total data) and the test set comprises the last 9275 measures (33%). The input time lags is as follows: 6 measures taken each 10 minutes, 6 taken each hour, 6 measures taken each 3 hours, 6 measures taken each 6 hours, 6 measures taken each 12 hours and 6 measures taken each 24 hours. Given this input time lag the goal is to predict the real value response.

The setting of the GP parameters for the tournament size was 8 and the mutation, reproduction and crossover probability were 0.1, 0.2 and 0.7, respectively. We set the inertia, cognitive and social weight of the PSO with the values 0.729, 1.49445 and 1.49445 respectively [31, 32]. The population size was 30 for the GP method and 50 for the PSO algorithm. The topology of ESN model consists of 36 input nodes and one output neuron. We present the results for different reservoir sizes in the range of 80 and 1000 neurons. Two important parameters of the model are the spectral radius and density of the reservoir weight matrix. we tested the model with spectral radius from 0.1 up to 0.9. We used sparse reservoir matrix, which have 70%, 80% and 90% of zero-values. The output parameters were computed using ridge regression algorithm with regularization parameter equal to 0.0001. The initial washout in the training and test procedure was of 30 samples.

We present in Table 1 the best accuracy obtained using the ESN model with different reservoir size, sparsity of the reservoir and spectral value of the reservoir matrix. In the case of the ESN model we used a network with 1000 reservoir neurons, 10% of non-zero connections and spectral radius equal to 0.1. Table 2 summarizes the accuracy of both models. The second row present the accuracy with the ESN with the following simple correction. If some predicted $y(t) < 0$

Table 1. Accuracy of the ESN model with the testing data set. First column corresponds to the number of reservoir neurons, second column corresponds to the sparsity of the reservoir matrix (percentage of non-zero values). Third column is the spectral radius. Last column correspond to the accuracy using the MSE using a scientific notation.

N_x	Sparsity	Spectral radius	MSE ($1.0e-3$)
80	10	0.4	0.7196
160	20	0.3	0.7039
350	10	0.4	0.7026
700	10	0.1	0.6871
1000	10	0.1	0.6826

Table 2. Accuracy of the ESN and the FNT model with the testing data set. The second row (ESN corrected) presents the performance of the model with the predicted values adjusted in $[0, 1]$.

Model	MSE
ESN	0.6826×10^{-3}
ESN (corrected)	0.6802×10^{-3}
FNT	0.76114×10^{-3}

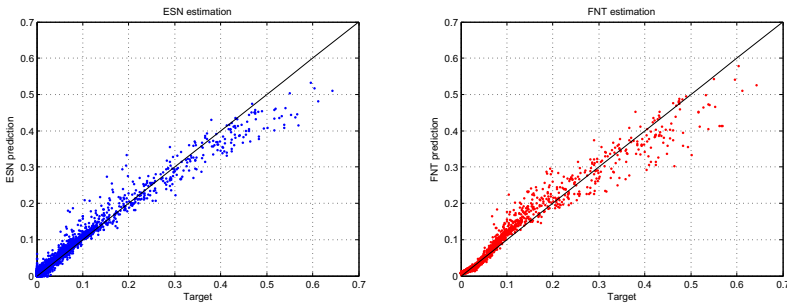


Fig. 1. Estimation of the ESN and FNT models on the test data set. The left figure shows the estimation of the ESN model with 1000 hidden units and sparsity equal to 0.1 and spectral radius 0.1. The right one shows the estimation of the FNT model. The black line is the identity function.

then we set the prediction to zero ($y(t) = 0$), this is given because it is known a priori that the output variable is positive. In the two graphics of the Figure 1 is presented the accuracy of both models. Both graphics present the target data in respect of the estimation data. In order to visualize the estimation quality in the graphics is presented also the identity function. Figure 2 presents examples

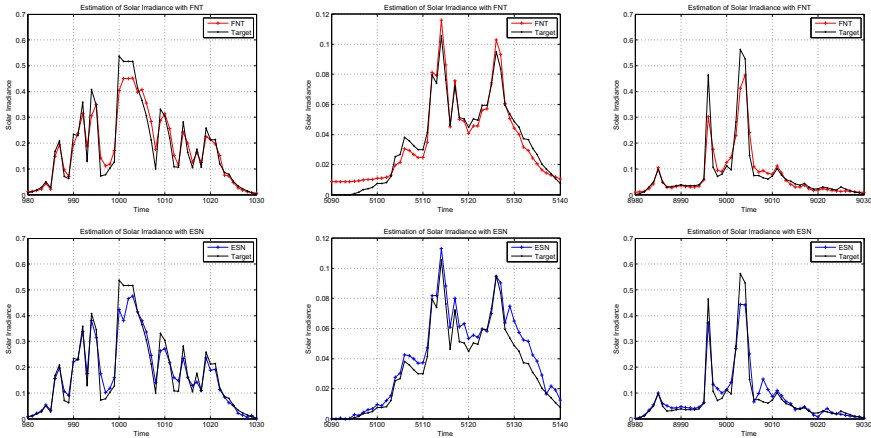


Fig. 2. Three examples of solar irradiance prediction. The first three figures illustrate the estimation of the FNT model. The three figures at the bottom show the estimation using the ESN model.

of the prediction using the FNT and ESN model on a window time of 50 of the test data set. In all cases, the left graphic corresponds to the ESN prediction and the right one corresponds to the FNT prediction.

5 Conclusions and Future Work

In the past decade were introduced two types of neural networks for solving temporal learning problems: the *Flexible Neural Tree (FNT)* and the *Echo State Network (ESN)*. The ESN is a recurrent neural network that uses the recurrences to model the temporal dependencies of the data. The model is characterized by a static recurrent structure that project the input data in a larger space. A memory-less function maps the projected space on the output space. The FNT is a particular multi-layer feedforward network. The model uses heuristic techniques for finding the parameters and its tree topology. There are several benchmarks in the Machine Learning literature that shown the success of both techniques for solving learning tasks. In this paper, we have studied the performance of these methods in a specific real data set about renewable energy. According to our results, the difference between both methods was not significant. Both learning models have been proven good accuracy for this problem. In future work, we will explore the capability of these two techniques using larger sliding windows time, for instance several days. Additionally, we will compare the results with other nonlinear time-series techniques.

Acknowledgments. This work was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070). Additionally, this article has been elaborated in the

framework of the project *New creative teams in priorities of scientific research*, reg. no. CZ.1.07/2.3.00/30.0055, supported by Operational Program Education for Competitiveness and co-financed by the European Social Fund and the state budget of the Czech Republic. This work was partially supported by the Grant of SGS No. SP2014/110, VŠB - Technical University of Ostrava, Czech Republic, and by the Bio-Inspired Methods: research, development and knowledge transfer project, reg. no. CZ.1.07/2.3.00/20.0073 funded by Operational Programme Education for Competitiveness, co-financed by ESF and state budget of the Czech Republic.

References

1. Chen, Y., Yang, B., Dong, J.: Nonlinear System Modelling Via Optimal Design of Neural Trees. *International Journal of Neural Systems* 14(02), 125–137 (2004)
2. Chen, Y., Yang, B., Dong, J., Abraham, A.: Time-series Forecasting using Flexible Neural Tree Model. *Inf. Sci.* 174(3-4), 219–235 (2005)
3. Jaeger, H.: The “echo state” approach to analysing and training recurrent neural networks. Technical Report 148, German National Research Center for Information Technology (2001)
4. Chen, Y., Abraham, A., Yang, B.: Hybrid Flexible Neural Tree based intrusion detection systems. *International Journal of Intelligent Systems* 22(4), 337–352 (2007)
5. Chen, Y., Abraham, A., Yang, B.: Feature selection and classification using Flexible Neural Tree. *Neurocomputing* 70(1-3), 305–313 (2006)
6. Chen, Y., Abraham, A., Yang, J.: Feature selection and intrusion detection using hybrid flexible neural tree. In: Wang, J., Liao, X.-F., Yi, Z. (eds.) *ISNN 2005*. LNCS, vol. 3498, pp. 439–444. Springer, Heidelberg (2005)
7. Chen, Y., Abraham, A., Yang, B.: Hybrid flexible neural-tree-based intrusion detection systems. *International Journal of Intelligent Systems* 22(4), 337–352 (2007)
8. Chen, Y., Yang, B., Dong, J.: Evolving Flexible Neural Networks Using ANT Programming and PSO Algorithm. In: Yin, F.-L., Wang, J., Guo, C. (eds.) *ISNN 2004*. LNCS, vol. 3173, pp. 211–216. Springer, Heidelberg (2004)
9. Kennedy, J., Eberhart, R.: Particle Swarm Optimization. In: *Proceedings of the IEEE International Conference on Neural Networks 1995*, vol. 4, pp. 1942–1948 (1995)
10. Shi, Y., Eberhart, R.C.: Parameter Selection in Particle Swarm Optimization. In: Porto, V.W., Waagen, D. (eds.) *EP 1998*. LNCS, vol. 1447, pp. 591–600. Springer, Heidelberg (1998)
11. Storn, R., Price, K.: Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *Journal of Global Optimization* 11(4), 341–359 (1997)
12. Price, K., Storn, R.M., Lampinen, J.A.: *Differential Evolution: A Practical Approach to Global Optimization*. Natural Computing Series. Springer-Verlag New York, Inc., Secaucus (2005)
13. Chen, Y., Peng, L., Abraham, A.: Exchange rate forecasting using flexible neural trees. In: Wang, J., Yi, Z., Žurada, J.M., Lu, B.-L., Yin, H. (eds.) *ISNN 2006*. LNCS, vol. 3973, pp. 518–523. Springer, Heidelberg (2006)
14. Shi, Y., Eberhart, R.: A modified Particle Swarm Optimizer. In: *The 1998 IEEE International Conference on Evolutionary Computation Proceedings of the 1998, IEEE World Congress on Computational Intelligence*, pp. 69–73 (1998)

15. Maass, W., Natschläger, T., Markram, H.: Real-time computing without stable states: a new framework for a neural computation based on perturbations. *Neural Computation*, 2531–2560 (November 2002)
16. Lukoševičius, M., Jaeger, H.: Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 127–149 (2009)
17. Schrauwen, B., Wardermann, M., Verstraeten, D., Steil, J.J., Stroobandt, D.: Improving Reservoirs using Intrinsic Plasticity. *Neurocomputing* 71, 1159–1171 (2007)
18. Steil, J.J.: Backpropagation-Decorrelation: online recurrent learning with $O(N)$ complexity. In: *Proceedings of IJCNN 2004*, vol. 1 (2004)
19. Xue, Y., Yang, L., Haykin, S.: Decoupled Echo State Networks with lateral inhibition. *Neural Networks* (3), 365–376 (2007)
20. Jaeger, H., Lukoševičius, M., Popovici, D., Siewert, U.: Optimization and applications of Echo State Networks with leaky-integrator neurons. *Neural Networks* (3), 335–352 (2007)
21. Gagliolo, M., Schmidhuber, J., Wierstra, D., Gomez, F.: Training Recurrent Networks by Evolino. *Neural Networks* 19, 757–779 (2007)
22. Basterrech, S., Rubino, G.: Echo State Queueing Network: a new Reservoir Computing learning tool. In: *IEEE Consumer Communications & Networking Conference (CCNC 2013)* (January 2013)
23. Verstraeten, D., Schrauwen, B., D’Haene, M., Stroobandt, D.: An experimental unification of reservoir computing methods. *Neural Networks* (3), 287–289 (2007)
24. Maass, W., Natschläger, T., Markram, H.: Computational models for generic cortical microcircuits. In: *Neuroscience Databases. A Practical Guide*, Boston, Usa, pp. 121–136. Kluwer Academic Publishers (June 2003)
25. Basterrech, S., Snášel, V.: Initializing Reservoirs With Exhibitory And Inhibitory Signals Using Unsupervised Learning Techniques. In: *International Symposium on Information and Communication Technology (SoICT)*, Danang, Viet Nam. ACM Digital Library (December 2013)
26. Basterrech, S., Fyfe, C., Rubino, G.: Self-Organizing Maps and Scale-Invariant Maps in Echo State Networks. In: *2011 11th International Conference on IEEE Intelligent Systems Design and Applications (ISDA)*, pp. 94–99 (November 2011)
27. Rodan, A., Tiño, P.: Minimum Complexity Echo State Network. *IEEE Transactions on Neural Networks*, 131–144 (2011)
28. Lukoševičius, M.: A practical guide to applying echo state networks. In: Montavon, G., Orr, G.B., Müller, K.-R. (eds.) *Neural Networks: Tricks of the Trade*, 2nd edn. LNCS, vol. 7700, pp. 659–686. Springer, Heidelberg (2012)
29. Prokop, L., Misak, S., Snasel, V., Platos, J., Kroemer, P.: Supervised learning of photovoltaic power plant output prediction models. *Neural Networks World* 23(4), 321–338 (2013)
30. Basterrech, S., Prokop, L., Buriánek, T., Misak, S.: Optimal Design of Neural Tree for Solar Power Prediction. In: *15th Scientific Conference Electronic Power Engineering*, Brno, Czech Republic (May 2014)
31. Eberhart, R.C., Shi, Y.: Comparing inertia weights and constriction factors in particle swarm optimization. In: *Proceedings of the 2000 Congress on Evolutionary Computation 2000*, vol. 1, pp. 84–88 (2000)
32. Clerc, M.: The swarm and the queen: towards a deterministic and adaptive particle swarm optimization. In: *Proceedings of the 1999 Congress on Evolutionary Computation, CEC 1999*, vol. 3, pp. 1951–1957 (1999)

A DOA Estimation Method for Wideband Signals with an Arbitrary Plane Array*

Jiaqi Zhen**, Qun Ding, and Bing Zhao

College of Electronic Engineering, Heilongjiang University, Harbin, 150080, China
zhenjiaqi2011@163.com

Abstract. The super-resolution direction finding for wideband signals usually requires preliminary DOA estimation, whether it is accurate or not will play an important part to the final result, in order to avoid the process, paper proposed a Direction Of Arrival (DOA) method for wideband signals without preliminary DOA estimation, focusing matrices are formed on transformation to the signal subspace of every frequency, the covariance matrix of the individually measured frequencies is gained, then method for narrowband signals was used to estimate the DOA. The algorithm decreases the computation and it is easy to implement, wherever, it can be adapted to an arbitrary plane array, computer simulations proved the effective performance of the method.

Keywords: direction of arrival, wideband signal, focusing matrix, signal subspace method.

1 Introduction

The spatial spectrum estimation super-resolution algorithms are widely used in radar, sonar and mobile communication in recent years, for example, multiple signal classification (MUSIC) [1] and estimation of signal parameters via rotational invariance techniques (ESPRIT)[2] are two representative methods of them, but they are only adapt to narrowband signals. Wideband signals can carry a large amount of information and has low probability of intercept, generally speaking, the Incoherent Signal-Subspace Method (ISSM)[3]and Coherent Signal Subspace Method (CSSM) [4,5] are commonly used. The latter is widely researched by their good performance, the basic idea is to focus the signal space of non-overlapping frequency band to the reference frequency, thus, a single frequency data covariance is obtained, then we can estimate their DOA by the methods which are adapt to narrowband signals. The scholars have proposed many methods building different focusing matrix, such as RSS[4,5], SST[6,7] and TCT[8,9], but the focus process of these methods has a large amount of calculation, it is a particularly prominent problem for DOA estimation of two-dimensional signals. In this paper, the focusing transformation matrices are achieved by signal subspace of each frequency, there is no focusing loss in the process, and the performance is compared with that of commonly TCT method.

* This work is supported by the National Natural Science Foundation of China (no. 61072072).

** Corresponding author.

2 Signal Model

It is shown in Fig.1, the setting of the source detection problem is stated as followed: assume that N far-field wideband signals impinge on M -element ($N < M$) arbitrary placed plane array from distinct directions $(\theta_1, \varphi_1), \dots, (\theta_N, \varphi_N)$, θ_i and φ_i is separately the azimuth and elevation of the i th signal, the coordinates of the m th sensor is (x_m, y_m) ($m = 1, 2, \dots, M$), the output of the i th sensor can be expressed:

$$x_m(t) = \sum_{i=1}^N s_i(t + \tau_{mi}) + n_m(t) \quad (m = 1, 2, \dots, M) \quad (1)$$

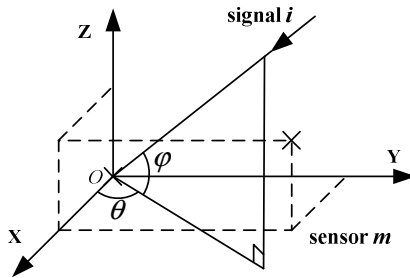


Fig. 1. Signal Model

Where $\tau_{mi} = \frac{x_m \cos \theta_i \cos \varphi_i + y_m \sin \theta_i \cos \varphi_i}{c}$, $n_m(t)$ is the temporally and spatially white Gaussian noise of m th sensor, suppose the observation time of data collection is ΔT , the initial sampling frequency of incident signal on each sensor is f_s , the sampling times is $K = \Delta T f_s$, formula (1) can be transformed by Discrete Fourier Transform(DFT):

$$X_m(f) = \sum_{i=1}^N S_i(f) \exp(-j2\pi f \tau_{mi}) + N_m(f) \quad (2)$$

Then the output of the array can be decomposed into some narrowband parts by filter bank, that is

$$X(f_i) = A(f_i, \theta_i, \varphi_i) S(f_i) + N(f_i) \quad i = 1, 2, \dots, J \quad (3)$$

3 Principle of the Method

According to the model of the wideband signal, the array covariance matrix of frequency f_i can be expressed as

$$\begin{aligned} R_{xx}(f_i) &= E[X(f_i)X^H(f_i)] \\ &= A(f_i)R_{ss}(f_i)A^H(f_i) + \sigma^2 I \end{aligned} \quad (4)$$

Where

$$\mathbf{R}_{ss}(f_i) = E[\mathbf{S}(f_i)\mathbf{S}^H(f_i)] \quad (5)$$

Set the reference frequency is f_0 , the focusing matrix can be constructed by the following equation

$$\mathbf{T}(f_i) = \frac{1}{\sqrt{J}}\mathbf{U}(f_0)\mathbf{U}^H(f_i) \quad (6)$$

Where J is the number of frequency samples, $\mathbf{U}(f_i)$ is the eigenvector of $\mathbf{R}_{xx}(f_i)$, it can be divided into the signal eigenvector and noise eigenvector, that is

$$\mathbf{U}(f_i) = [\mathbf{U}_{ss}(f_i) \quad \mathbf{U}_N(f_i)] \quad (7)$$

Where $\mathbf{U}_{ss}(f_i)$ is $M \times N$ -dimensional matrix, it corresponds to the N larger signal eigenvalues; $\mathbf{U}_N(f_i)$ is a $M \times (M - N)$ -dimensional matrix, it corresponds to the $M - N$ smaller noise eigenvalues, as the focusing matrix is constructed by $\mathbf{U}(f_i)$, so we do not need to know the number of signals, define

$$\mathbf{Y}(f_i) = \mathbf{T}(f_i)\mathbf{X}(f_i) \quad (8)$$

The covariance matrix is solved by the equation above, that is

$$\begin{aligned} \mathbf{R}_{YY}(f_i) &= E[\mathbf{Y}(f_i)\mathbf{Y}^H(f_i)] \\ &= \mathbf{T}(f_i)\mathbf{A}(f_i)\mathbf{R}_{ss}(f_i)\mathbf{A}^H(f_i)\mathbf{T}^H(f_i) + \frac{1}{J}\sigma_n^2\mathbf{I} \end{aligned} \quad (9)$$

As

$$\begin{aligned} &\mathbf{T}(f_i)\mathbf{A}(f_i) \\ &= \frac{1}{\sqrt{J}}[\mathbf{U}_{ss}(f_0) \quad \mathbf{U}_N(f_0)] \begin{bmatrix} \mathbf{U}_{ss}^H(f_i) \\ \mathbf{U}_N^H(f_i) \end{bmatrix} \mathbf{A}(f_i) \\ &= \frac{1}{\sqrt{J}}[\mathbf{U}_{ss}(f_0) \quad \mathbf{U}_N(f_0)] \begin{bmatrix} \mathbf{U}_{ss}^H(f_i)\mathbf{A}(f_i) \\ \mathbf{0} \end{bmatrix} \\ &= \frac{1}{\sqrt{J}}\mathbf{U}_{ss}(f_0)\mathbf{U}_{ss}^H(f_i)\mathbf{A}(f_i) \end{aligned} \quad (10)$$

Take the formula (10) into (9), we can obtain

$$\mathbf{R}_{YY}(f_i) = \frac{1}{J}\mathbf{U}_{ss}(f_0)\tilde{\mathbf{R}}(f_i)\mathbf{U}_{ss}^H(f_0) + \frac{1}{J}\sigma_n^2\mathbf{I} \quad (11)$$

where

$$\tilde{\mathbf{R}}(f_i) = \mathbf{U}_{ss}^H(f_i)\mathbf{A}(f_i)\mathbf{R}_{ss}(f_i)\mathbf{A}^H(f_i)\mathbf{U}_{ss}(f_i) \quad (12)$$

The universal focused correlation matrix is acquired by adding all of the $\mathbf{R}_{YY}(f_i)$ of every frequency, that is

$$\begin{aligned}
\mathbf{R}_{sum} &= \sum_{i=1}^J \mathbf{R}_{YY}(f_i) \\
&= \frac{1}{J} \mathbf{U}_{SS}(f_0) \left(\sum_{i=1}^J \tilde{\mathbf{R}}(f_i) \right) \mathbf{U}_{SS}^H(f_0) + \sigma_n^2 \mathbf{I} \\
&= \mathbf{U}_{SS}(f_0) \mathbf{R}_S \mathbf{U}_{SS}^H(f_0) + \sigma_n^2 \mathbf{I}
\end{aligned} \tag{13}$$

where

$$\mathbf{R}_S = \frac{1}{J} \sum_{i=1}^J \tilde{\mathbf{R}}(f_i) = \frac{1}{J} \sum_{i=1}^J \mathbf{U}_{SS}^H(f_i) \mathbf{A}(f_i) \mathbf{R}_{SS}(f_i) \mathbf{A}^H(f_i) \mathbf{U}_{SS}(f_i) \tag{14}$$

It can be proved that [10], matrix \mathbf{R}_{sum} is always full rank, that is

$$\text{rank}(\mathbf{R}_{sum}) = N \tag{15}$$

We add all of the covariance matrices of every frequency by equation (14), so the covariance matrix of different frequency can be focused at the reference frequency f_0 , here, as every column of matrix $\mathbf{U}_{SS}(f_0)$ and $\mathbf{A}(f_0)$ span the same subspace, $\mathbf{U}_{SS}(f_0)$ can be used for focusing instead of $\mathbf{A}(f_0)$, the course of preprocessing for angles is avoided. Also, because the process of focusing doesn't affect the characteristics of Gaussian white noise, that of the autocorrelation matrix is unchanged, conventional narrowband MUSIC algorithm can be used for the subsequent DOA estimation. There are no specific requirements to the form of the array for the method of the paper, it is adapt to arbitrary plane array(APA), so it can be called APA method.

Suppose the frequency samples is J , comparing with the traditional TCT algorithm and SST algorithm which are usually used for direction finding to wideband signals, the process of the focusing makes use of the eigenvectors of the covariance matrices of every frequency which has no noise, but that of the paper directly makes use of the eigenvectors of the covariance matrices of every frequency, the computation of the TCT algorithm is MJ times more than that of the paper; wherever, that of the SST algorithm makes use of the singular vectors of the covariance matrices of every frequency, its computation is nearly the same, but as it usually needs the preprocessing by the traditional method of beamforming before the proposed method, the process not only adds the complex degree of computation, but also may leads to the larger error for the final estimating result.

The proposed method of the paper is summarized as follows:

1. Apply a DFT to the array output to sample the spectrum of data, then use equation (4) to solve the covariance matrix;
2. Determine the reference frequency f_0 ;
3. Perform the eigen-decomposition to the covariance matrix above, calculate focusing matrix by formula (6);
4. Evaluate $\mathbf{R}_{YY}(f_i)$ of every frequency by formula (9);
5. Estimate \mathbf{R}_{sum} by formula (13), then apply MUSIC or other high-resolution narrowband spectral estimation method to find the DOA.

4 Analytical Study of the Experimental Results

In order to verify the effective of the APA method, three simulations are presented with matlab below, consider some wideband chirp signals impinge on 8 arbitrary placed plane array, their coordinates are (0, 0), (-0.13, 0.13), (-0.056, 0.11), (-0.12, 0.061), (-0.16, -0.053), (0.049, 0.09), (0.17, 0.07), (0.055, -0.038), it is in meters, the center frequency of the signals is 4GHz, the width of the band is 30% of the center frequency, APA, TCT and SST algorithm are separately used for the simulations, and comparing with their spatial spectrum figures、 resolution probability and angle measurement accuracy, the center frequency of the signals is selected as the reference frequency. The computer is Pentium dual-core processor, clock is 2.5GHz, 3.25MHz memory.

4.1 Simulation 1 Spatial Spectrum Figure

In the first simulation, three wideband coherent signals with the same power impinge on the array from directions $(55^\circ, 36^\circ)$ 、 $(164^\circ, 58^\circ)$ 、 $(276^\circ, 80^\circ)$, the snapshots of every frequency is 100, searching steps is 0.5° , Signal-To-Noise ratio (SNR) is 6dB, the spatial spectrum are shown as Figure.2~Figure.4 below.

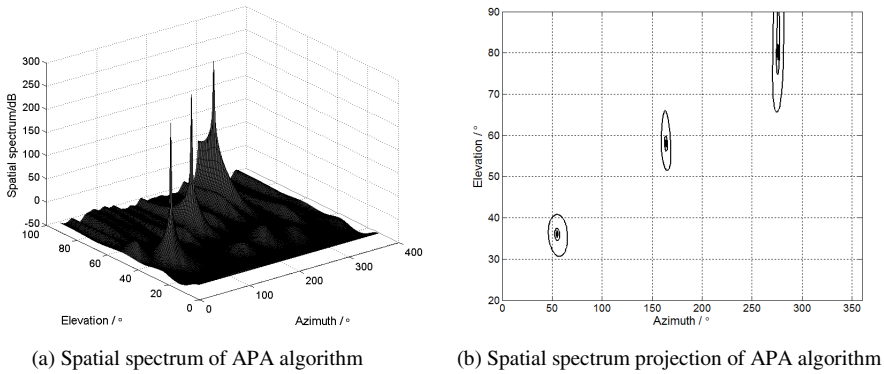


Fig. 2. Spatial spectrum figure of APA algorithm

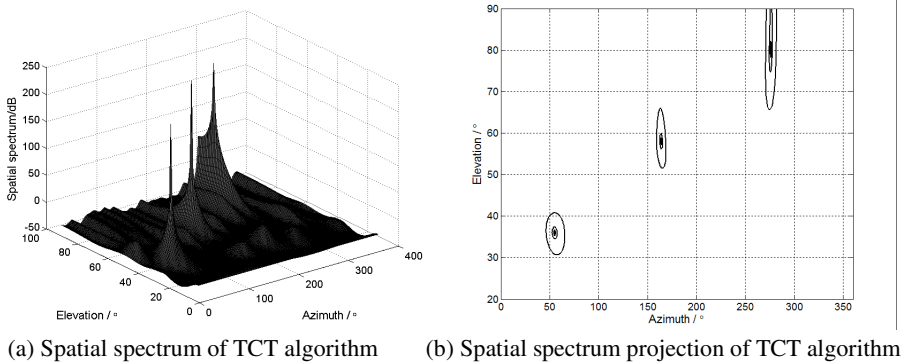
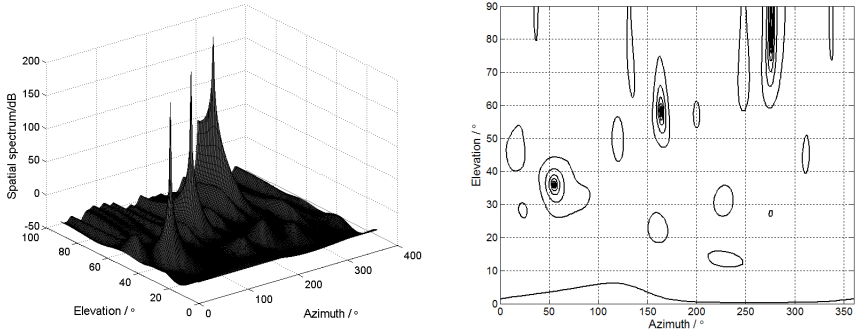


Fig. 3. Spatial spectrum figure of TCT algorithm



(a) Spatial spectrum of SST algorithm (b) Spatial spectrum projection of SST algorithm

Fig. 4. Spatial spectrum figure of SST algorithm

It is seen from Figure.2~Figure.4, all of the algorithms can estimate the directions of the coherent signals, and the spatial spectrum figures of APA and TCT algorithm are very close, but that of the SST algorithm is obviously less sharper than the above two, meanwhile, it has some more and small fake ones besides the actual peaks.

4.2 Simulation 2 Resolution Capability

Consider two far-filed signals with the same power impinge on the array, the direction (θ, φ) is defined as ϑ , in the field of the super-resolution direction finding technology, generally speaking, resolving power boundary is generally defined that the peak of mean value of angles of two signals is equal to the mean peak value of two signals for spatial spectrum algorithm [11], that is

$$P(\vartheta_m) = P_{\text{peak}} \tag{16}$$

where, $\vartheta_m = (\vartheta_1 + \vartheta_2) / 2$ is the mean value of two signals; $P(\vartheta_m)$ is the peak of that; P_{peak} is mean peak value of two signals, and the following equality is satisfied:

$$P_{\text{peak}} = \frac{1}{2} [P(\vartheta_1) + P(\vartheta_2)] \tag{17}$$

In the course of searching spectrum peaks, if there is hollow between two spectrum functions of the signals, namely left of the equation(18) is less than the right, it is thought that the two signals can be resolved; if not, they can't be resolved. Resolution capability is defined as:

$$\gamma(\Delta) = 1 - \frac{\left| F\left(\frac{\Delta}{2}\right) \right|^2}{1 - |F(\Delta)|^2} \cdot \{1 - |F(\Delta)| \cos[\varphi_{F(\Delta)} - 2\varphi_{F(\Delta/2)}]\} \tag{18}$$

$$F(\Delta) = \frac{1}{M} \sum_{i=1}^M \exp\{-j \frac{2\pi}{\lambda} (x_i \cos \phi + y_i \sin \phi) \times (\cos \theta_2 - \cos \theta_1)\} \tag{19}$$

$\gamma(\Delta)$ is used for measuring whether the signals can be resolved, and it is defined as the measuring standard of resolution capability of two incident signals, the size of $\gamma(\Delta)$ is represented as degree of concavity between two spectrum peaks, the larger $\gamma(\Delta)$ is, the greater the degree of concavity is, the more power the resolution capability for two incident signal with space Δ is.

Consider both of azimuths of the two far-field wideband signals with the same power are 50° , their elevations are taken 75° as point of symmetry, the angle space is ranging from 0.5° to 15° , the step size is 0.5° , 200 times Monte-Carlo trials have run for each Δ , the average of them is regarded as the measure result for this Δ , other conditions are the same with simulation 1, here, we present the simulation result for resolution capability of three methods with different angle spaces, it is shown in Fig.5.

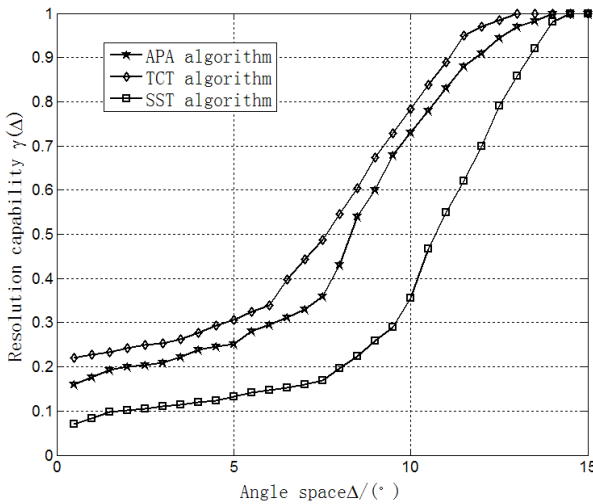


Fig. 5. Effect of algorithm resolution with the angle space

It is seen from Fig.5, all of the resolution capabilities of the three methods is not high enough when the angle space is small, and they change slowly, as the angle space growing, their resolution capabilities are all improving. In which TCT is the best, APA is the second, SST is the third.

4.3 Simulation 3 Angle Measurement Accuracy

Consider two far-field wideband signals with the same power arriving at the sensors from $(40^\circ, 70^\circ)$, $(50^\circ, 80^\circ)$, SNR varies from -5dB to 15dB , step size is 1dB , other conditions are the same with simulation 1, 200 times Monte Carlo trials have run for each SNR, the average of them is regarded as the measure result for this SNR. In

order to describe the angle measurement accuracy, Root Mean Squared Error (RMSE) of two-dimensional angle measurement is defined as:

$$RMSE = \sum_{i=1}^2 \sqrt{(\hat{\phi}_i - \phi_i)^2 + (\hat{\theta}_i - \theta_i)^2} \quad (i = 1, 2) \quad (20)$$

ϕ_i and θ_i ($i=1,2$) are separately the actual values of azimuth and elevation of the i th signal, $\hat{\phi}_i$ and $\hat{\theta}_i$ are separately the values estimated, Fig.6 shows the RMSE of the three methods with SNR.

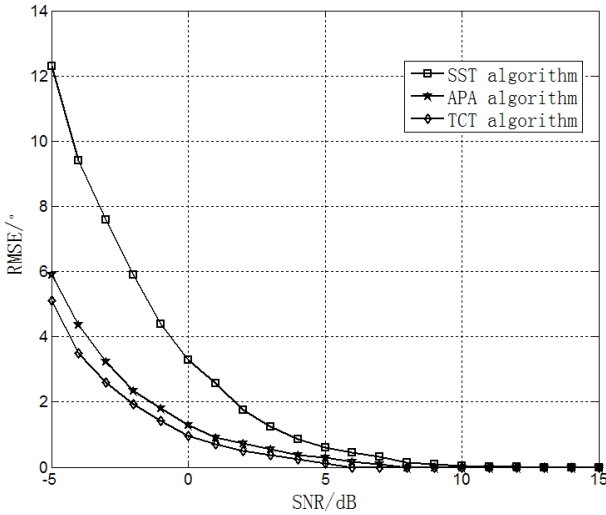


Fig. 6. The RMSE of three methods with SNR

Define the sensors is M , number of frequency is J , the number of samples of every frequency is Kp , according to the analysis, JM^2Kp times of multiplication, $JM^3 - JM^2 + J$ times of addition, MJ times of subtraction and once division is needed in the process of focusing; APA needs JM^2Kp times of multiplication, $JM^3 - JM^2 + J$ times of addition and once division, although the computation of the process of SST and APA is nearly the same, SST needs the preprocessing of initial estimate to the DOA, it also spends some time, associating with the specific methods, so there is no advantage in spending time comparing with APA, but both of them have less time of removing noise part comparing with TCT. Table 1 has shown the computation times and average time for estimating final DOA (does not include the preprocessing of initial estimate to the DOA).

It is seen from table 1, the computation time of APA the paper proposed is less than that of TCT, but a little more than that of SST; and we know from Fig.6, the measurement accuracy of APA is higher than that of SST, and not obviously difference with TCT.

Table 1. Average computation time of three methods

Method	Example	Computation time(s)
TCT	$JM^2Kp+JM^3-JM^2+J+MJ+1$	0.7822
SST	$JM^2Kp+JM^3-JM^2+J+1$	0.7328
APA	$JM^2Kp+JM^3-JM^2+J+1$	0.7332

5 Conclusion

The paper proposed a new method of focusing to the problem of large amount of computation for traditional wideband direction finding methods, it overcomes the shortcoming of need for pre-estimate to the direction, the focusing matrix is directly built by the receiving signals, reducing the computational complexity and improving the efficiency, simulation results have shown that its computation complexity is less than classic TCT, and has no significant difference in the estimation accuracy. Besides, the method has low demand for the position of the plane array, it adapts to arbitrary plane array for two-dimensional DOA estimation, but it also belongs to subspace focusing direction finding methods, selection of the proper reference frequency is needed, how to implement the process is worthy of further study, it is very important for the realization in the project.

References

- Schmidt, R.O.: Multiple emitter location and signal parameter estimation. J. IEEE Trans. on Antennas and Propagation 34, 276–280 (1986)
- Roy, R., Kailath, T.: ESPRIT-estimation of signal parameters via rotational invariance techniques. J. IEEE Trans. on Acoustics, Speech and Signal Processing 37, 984–995 (1989)
- Su, G., Morf, M.: Signal subspace approach for multiple wideband emitter location. J. IEEE Transactions on Acoustics, Speech and Signal Processing 31, 1502–1522 (1983)
- Salman, N., Ghogho, M., Andrew, H.: On the Joint Estimation of the RSS-Based Location and Path-loss Exponent. J. IEEE Wireless Communications Letters 1, 34–37 (2012)
- Hung, H., Kaveh, M.: Focussing matrices for coherent signal-subspace processing. J. IEEE Transactions on Acoustics, Speech, and Signal Processing 36, 1272–1281 (1988)
- Qasim, Z.A., Yang, L.L.: Reduced-Rank Adaptive Multiuser Detection in Hybrid Direct-Sequence Time-Hopping Ultrawide Bandwidth Systems. J. IEEE Transactions on Wireless Communications 9, 156–167 (2010)
- Doron, M.A., Weiss, A.J.: On focusing matrices for wideband array processing. J. IEEE Transactions on Signal Processing 40, 1292–1302 (1992)
- Valaee, S., Kabal, P.: Wideband array processing using a two-sided correlation transformation. J. IEEE Transactions on Signal Processing 43, 160–172 (1995)
- Luo, P., Liu, K.H., Yu, J.X.: Novel DOA estimation method for coherent wideband LFM signals. J. Journal on Communications 33, 122–129 (2012)

10. Pasadena, P., Vaidyanathan, P.: A Novel Autofocusing Approach for Estimating Directions-of-Arrival of Wideband Signals. In: Proceedings of the Forty-Third Asilomar Conference on Signals, Systems and Computers, pp. 1663–1667. IEEE Press, Pacific Grove (2009)
11. Wang, Y.L., Chen, H., Peng, Y.N.: Spatial spectrum estimation theory and algorithm. Tsinghua University, Beijing (2004)

An e-Learning System Based on EGL and Web 2.0

Xiaomei Li¹, Zhaozhe Ma², and Bo Song^{3,*}

¹ Research and Training Center for Basic Education, Shenyang Normal University,
Liaoning Shenyang 110034, China

² Ming Shan Teacher Training School, Liaoning Benxi 117000, China

³ College of Software, Shenyang Normal University,
Liaoning Shenyang 110034, China
songbo63@aliyun.com

Abstract. Aiming at the existing problems of e-Learning system application architecture, in the analysis of the relationship between EGL and Web 2.0 technologies, an application architecture of e-Learning system based on EGL and integrating Web 2.0 technology is proposed in this paper. Through the process of design and implementation of an e-Learning system, the key feature of the architecture is demonstrated – developers can focus on the business issues what code handle without caring for software technical details. The architecture is simple, easy to use and across languages, frameworks and runtime platforms. In addition, it can reduce the cost during the development stage of application and effectively improve the real-time requirements and human- computer interaction experience of e-Learning system.

Keywords: e-Learning, EGL, Web 2.0, architecture.

1 Introduction

With the development of Web 2.0 technology, the rich client in RIA (Rich Internet Application) is rapidly replacing the thin client in B/S [1]. Because the e-Learning system can provide richer end-user experience, it has been adopted by more and more developers based on RIA architecture. For e-Learning system, the main benefit of Ajax is a greatly improved user experience. Although JavaScript and DHTML – the technical foundations of Ajax – have been available for years, most programmers ignored them because they were difficult to master. Although most of the Ajax frameworks available today simplify development work, you still need a good grasp of the technology stack. So, if you're planning to use Ajax to improve only your application's user experience – if you're not also using it as a strategic advantage for your business – it may be unwise to spend a lot of money and time on the technology [2]. ORM (Object-Relational Mapping) is a kind of technology which can solve the problem of impedance mismatch between Object-Oriented Programming and RDB (Relational Database). EJB 3, Hibernate and Oracle TopLink are the effective solutions

* Corresponding author.

to implement ORM [3] [4], but the implementation of ORM is time-consuming compared with JDBC [5]. Although the foregoing ORM tools provide convenience for operating RDB as object, the cost and complexity of e-Learning system are increased, and the real-time requirement of e-Learning system is reduced [6].

Aiming at the above-mentioned problems, an application architecture of e-Learning system based on EGL (Enterprise Generation Language) and Web 2.0 technology is proposed in this paper. Using the features of a cross platform and across application of EGL, the architecture can develop Web 2.0 application applied to browser-side and Java application running on server-side only with EGL. Developers do not need to master the two language – JavaScript and Java at the same time. The key feature of the architecture – developers can focus on the business issues what code handle without caring for software technical details – is implemented by using existing platform and technology instead of replacing them and improve the real-time requirements and human- computer interaction experience of e-Learning system effectively.

2 EGL and Web 2.0

EGL is a high-level language that lets developers create business software without requiring that they have a detailed knowledge of runtime technologies or that they be familiar with object-oriented programming. The language is architected to reflect patterns that are common to different kinds of business software, and the language hides many details that are platform specific. EGL also helps a company retain developers who are knowledgeable in business processes, even if those developers lack the time needed to stay current with technical change. And the relative simplicity of the language helps traditional developers become accustomed to the latest technologies.

The rational products that support EGL are based on Eclipse, which is the IDE (integrated development environment). EDT (EGL Development Tools) includes the core language packages – EGL SDK, and corresponding IDE [7]. In the EDT environment, EGL is used in a development process that has defined steps, from coding a source to generating an output (Java, or JavaScript) to preparing and deploying that output. EDT support the development of Web 2.0 application and its deployment. Terminal user access the Web page (include HTML and JavaScript) generated by EGL code and the browser is responsible for download them to client-side. On the client-side, JavaScript generated by EGL code will interpreter in the browser and demonstrate the corresponding interface. Then JavaScript code generated by corresponding EGL statement is responsible for calling the Web Service or REST Service deployed on the server. Java EE container on the server-side is responsible for receiving the request from client-side and returning it to the browser, and then JavaScript application generated by EGL on the client-side will demonstrate it to the terminal user. As is shown in the compilation process of EGL, EGL itself does not run directly and it will be compiled and executed by the compiler of target language in building and generating target language on corresponding platform. And in the process of language generation, several of different platforms and existing technology

can be integrated and made full use of. For example, in the browser-side, EDT made full use of Dojo framework to support the development of Web 2.0 and JavaScript generated by EGL encapsulated the Dojo [8]; in the server-side, Java code generated by EGL encapsulated database access using JDBC. EGL is not to replace existing technology and not to unify to develop new language and it is to maximize the use of mature technology as well as the supplement and extension of existing technology.

3 Implementation of the e-Learning System

As Fig.1 showed, a kind of hierarchical and extensible e-Learning system is proposed in this paper. After a Web server transmits the RUI (Rich User Interface) application to the user's browser, subsequent interaction with the server occurs only if the browser-based code accesses a service, which is a unit of logic that is more-or-less independent of any other unit of logic. The RUI application can access any number of discrete services. You automatically deploy the RUI application with the EGL RUI Proxy, which is EGL runtime code that handles the communication between the RUI application and the accessed services. The distinction among the tiers gives you a way to think about the different kinds of processing that occurs at run time.

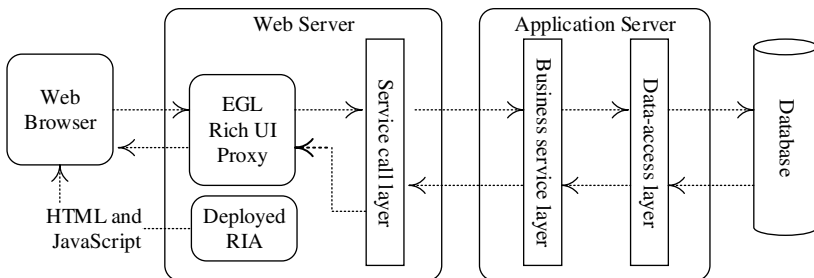


Fig. 1. The architecture of the e-Learning system

The architecture proposed by this paper makes full use of the characteristic of EGL technology and obtains further deputation and encapsulation, which enable the framework more suitable for the design and development of specific e-Learning system. The EGL RUI Proxy is runtime software that is installed with your RUI application if the deployment target is IBM WebSphere Application Server or Apache Tomcat. The EGL RUI Proxy handles communication between the application and any services that are accessed by the application [9].

3.1 Database Access Layer

The basic idea of a relational database is that data is stored in persistent tables. Each table column represents a discrete unit of data and each row represents a collection of such data and is equivalent to a file record. An EGL record can be the basic of a

variable used as the source or target of an I/O operation [9]. We can interact with a relational database as follow: define a record whose stereotype is *SQLRecord*, create a variable based on that record and use the variable as an I/O object in different data-access statements. Usually, one or more columns in a database table can be primary keys, which mean that the values in those columns are unique to a given row. For example, here is a Record about lessons of the e-Learning system:

```
record Lesson type Entity {
    @table{name = "USER.LESSON"}}
    lesson_id int { @id, @GeneratedValue, @Column{name="lesson_id"}};
    lessonname string(100) { @Column{name="lessonname"}};
    lessonintro string(255){ @Column{name="lessoninfo"}};
end
```

The *lesson_id* column is an identity column, which means that the database will place a unique value into that column whenever the user creates a record. Each value is one more than the last. To generate the Java code that is appropriate for the SQL operation, the EDT Java generator uses the Table, ID, Generated Value, and Column annotations. Access to relational database is by way of SQL. To retrieve a row, we can assign a value to the record field associated with the key column and then issue a get SQL statement to read data from the table in database. The code of function *getAllLessons* using get SQL statement is shown below:

```
function getAllLessons () returns (Lesson [])
    Lessons Lesson[];
    try
        get lessons from MyService with #sql {
            select lesson_id, lessonname, lessonintro from LESSON
        };
        onException (ex sqlException)
    end
    return (Lessons);
end
```

The following code declares a related record variable in the client-side handler. And then we can use the variable to access the database.

```
myLesson Lesson [ ];
```

In addition, before we generate EGL code, we should configure a build descriptor, which is a build part that guides the generation process and references other definitions. The resource associations can associate the logical file name with a physical file on each target platform where we intend to run the code.

3.2 Business Service Layer

Services can include new logic and can expose the data returned from other services and from called programs. EGL language offers end-to-end processing: developers can write the user interface, the service logic, and, if necessary, new backend programs.

A RUI application invokes services asynchronously, which means that the user can still interact with the user interface while the RUI application is waiting for the service to respond. However, if the user needs the information to continue a task, we can disable widgets and present a simple animation until the service responds. The runtime technology ensures that the invocation occurs as soon as a message arrives from the service. The process for invoking a service often requires an EGL interface part, which describes the data that can pass between the application and service. The Interface part tells the names of the service operations and, for a given operation, the kinds of data that the application exchanges with the service [9].

In many cases, before invoking the service, we need declare the variable based on the interface part in the client-side handler. Here is a declaration of *MyService*:

```
MyService SQLDataSource?;
dedicatedServiceBinding HTTPProxy;
```

In addition to the service variable declaration in the client-side handler, the service variable in the service is also needed.

```
MyService SQLDataSource? {
    @Resource{uri="binding:eLearnDerby" }
};
```

The service variable declaration specifies the location which identifies the protocol that formats a message at the start of transmission and unformatted the message at the end of transmission. Web Service is a facility to let developers create logic that receives or sends messages over HTTP. The “*eLearnDerby*” is a connection to database. The called service can be available by binding the database connection. This information can all be included in a Web Services Description Language (WSDL) file. A service contains public functions that can be accessed from other code and can include private functions and global variables, but those functions and variables are solely for use by functions that are within the service. At run time, the service is stateless, which means that the internal logic never relies on data from a previous invocation.

3.3 Rich UI Interface

In the RUI design of e-Learning system, the basic design idea is to refresh RUI widget as a unit. That is, the whole page is divided into several widgets and each widget varies independently. When retrieving data by calling backend service, EDT need to refresh

the front page and the grain size of refreshed widgets should be as small as possible, so that it can reduce the throughput and response time. In order to realize the design, a global access control point of the refresh widgets is needed. Because EGL access service using asynchronously calling, each calling will construct a callback function instance. The instance is responsible for refresh the user interface after returning the calling results. If the instance accesses the widgets of the page, it must have a reference to the widgets.

To create an EGL RUI application, a RUI handler is need primarily. The handler holds the EGL logic to add widgets to an initial DOM tree and to respond to events such as a user's click of a button. Primarily, an EGL RUI handler named *MainHandler* is created as a whole to call other handlers that can be designed as the sub modules of the e-Learning system by user's click of buttons. The buttons should be added an event named *showcall* as is shown below:

```
onClick ::= showcall;
```

Then we can configure the code so that the event handler responds to the event that is internal to the code. And the function *showcall* runs as soon as the user clicks the buttons. Such an event might be receipt of a message that was returned from a service. In the *MainHandler*, we can write the event handler and call other handlers by using a case statement. The code is shown below:

```
function showcall (event Event in)
  button DojoButton=event.widget;
  BoxContent.children = [ ];
  case (button.text)
    when ("HomePage") BoxContent.appendChild(new LoginHandler{ }.ui);
    when ("e-Learn") BoxContent.appendChild(new Learning{ }.ui);
    ...
  end
end
```

The button widgets integrated the Dojo widgets and were placed into Box and GridLayout widget of the page. The page is divided into several sections by Layout widgets and the Box named *BoxContent* belongs to Layout widget. The sub modules will be the children of *BoxContent* and compose the single application by embedding multiple RUI handlers. However, by saying “embedded handlers” we do not mean to say that we physically embed one handler in another. Instead, one handler – an EGL part that present the user interface – declares a variable used to access the functions and widgets in a second handler. For example, we can not only declare a variable that provides access to the handler *LoginHandler* as shown below, but also access it directly using “new” keyword as shown in the case statement of the function *showcall*.

```
myLoginHandler LoginHandler { };
```

A reasonable practice is to use embedded handlers for service invocation and for other business processing that lacks a user interface. If the embedded handler has an on- construction function, the function runs when the declaration for the related variable runs. The core module of the e-Learning system is Learning Module which is included in the handler named *Learning {}*. It encapsulates the references of some RUI widgets such as buttons including events, dataGrid displaying the information of lessons of the e-Learning system from database. These events can be executed by calling service from business service layer. Each service calling will construct a callback function instance and retrieve the references of the needed RUI widgets. By the RUI, service calling layer can be responsible for calling the service provided by the Business service layer on the server-side and then return the results to client logic layer and client presentation. That is to say, service calling layer decouple the front logic from the backend logic and use call statement to call the created *MyService* of each service. The access of Backend service becomes simpler and the code becomes easier to maintain. Then Take function *readFromTable* for example to demonstrate the implementation principle of service calling layer. The widgets of client-side include UI controls and each widget can indicate screen events such as “*onClick*” to call the code of service layer.

```
function readFromTable (event Event in)
  call MyService.getAllLessons( )
  using dedicatedServiceBinding returning to mycallback
  onException serviceExceptionHandler;
end
```

The function *readFromTable* is responsible for retrieving data and displaying the data. As is show in the code of the function, the screen event “*onClick*” will call the service of *MyService* using *dedicatedServiceBinding*. Resource Binding is one of the outstanding characteristics of EGL language. This simply means it is a description about how to connect to the database and how to invoke the service. We can maintain the binding in the deployment descriptor files of EGL and the binding can be seen the extending of the application logic. When we develop and deploy the application, the deployment descriptor files will provide specific details of connection and calling service. If errors occur in the process of calling service, exception handler *serviceExceptionHandler* will be called. After the screen event “*onClick*” retrieve the data from the database by service calling layer, and then it will use the function *mycallback* to process it. In this case, the DataGrid widget named *myLesson* is responsible for displaying the data. The code of *mycallback* is shown as below:

```
function mycallback (retResult Lesson[] in)
  myLesson = retResult;
  myLesson_ui.data = myLesson as any [];
end
```

We can access a function or property in an embedded widget by extending the dot syntax. For example, the above-mentioned statement retrieves the displayed data of the DataGrid widget named *myLesson*. In the Learning handler, users can not only retrieve the information of lessons from database, but also add, delete or edit the lessons into the RUI. These events can also be realized by buttons and they can be displayed by function *showDialog* and also be hidid by function *hideDialog*.

```
function showDialog (event Event in)
    Dialogcontent.children = [info,buttonBar];
    dialog.showDialog();
end
function hideDialog (event Event in)
    dialog.hideDialog();
end
```

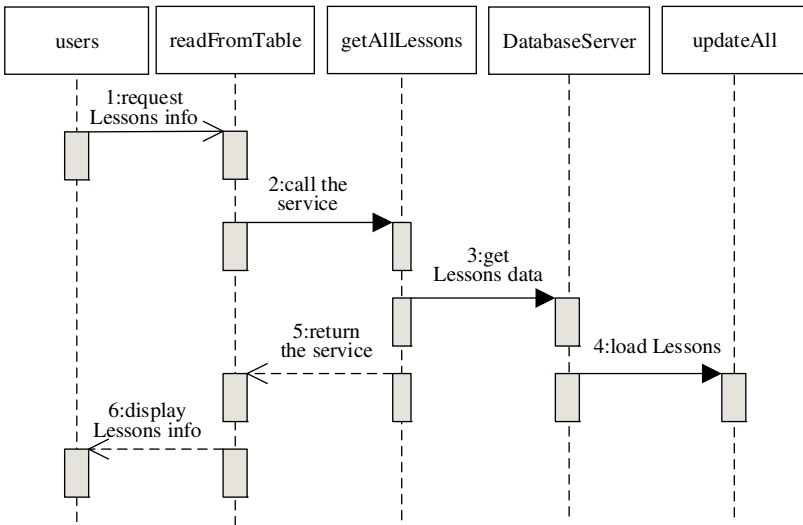


Fig. 2. The call flows between client-side and server-side

As is shown in Fig.2, the function *readFromTable* creates the screen event “*OnClick*” and prepares to call the service with call statement. The function *getAllLessons* is a function of the service which invokes access the database with SQL statement. The function *onException* is responsible for an error value in the process of data access. The function *updateAll* is a callback function and is responsible for handling the data returned from the service. The function *readFromTable* and *updateAll* are all in the handler of client-side as Rich UI and the function *getAllLessons* in the service file locate in business service layer in server-side. The focus of business service logic layer is the development of business rules and the implementation of business flow which is related to business requirements.

Any technique for working with a service is a variation on what we have shown here: create a variable, bind it to a service client binding, and access a function by way of the variable. Remote Services are the standard way to communicate with the Web application server from the user's browser. This is done using a standard Ajax, essentially an HTTP POST. The server-side code for that Ajax call is found within the generated JavaScript application. This is code that executes a Web application server and this is where we tie user actions at the browser into business logic and then for data persistence storage.

4 Conclusion

In this paper, it is simulated that the client-side submits request to the Web server-side in the EDT development environment to propose application architecture of e-Learning system based on EGL and Web 2.0. The testing scenario is information query in the handler of client-side, users can create and connect the database connection and retrieve the data in database through access the service. The aggregate average response time (AART) of the system is shown as Fig.3.

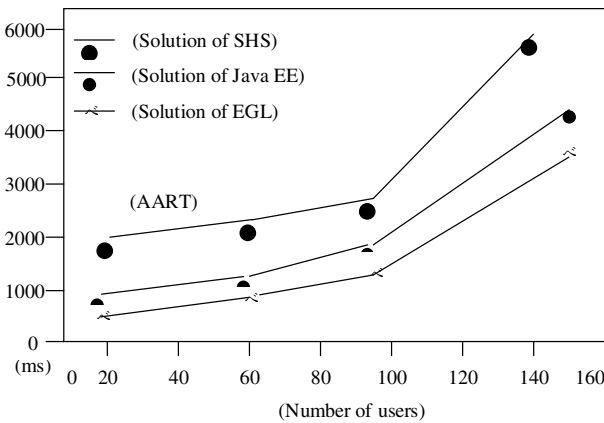


Fig. 3. AART curve

Analysis of AART curve in Fig.3, compared with the e-Learning system solution based on open source framework—Struts, Hibernate and Spring (SHS) [4], the system response time of the solution of Java EE increases in accordance with the linear way until it reaches about 100 users [6]. After that time, the increase of the curve becomes more intense. By analyzing the change trend of the two curves, we can draw the conclusion: the maximum number of users of the solution of Java EE presents is 100 and the application is suitable for small e-Learning system. However, the solution can significantly reduce the cost and complexity of e-Learning system and improve the e-Learning system interactive experience and real time requirements. Because EGL language can be executed after generating Java code, the changes of the curve of the

solution of EGL are similar to Java EE's and only its speed is slightly slow. It implements the loose coupling between business logic and access control, that is, choices related to a user interface, and choices related to the data in persistent, are handled separately. EGL is general because the technologies are so varied, but we will follow up by nothing how the separation applies to EGL Rich UI. In the development of EGL Rich UI application, the MVC pattern is often used to realize above-mentioned features. Therefore, the primary benefit of the application architecture of e-Learning system based on EGL is simplicity. The separation of Model and View also allows for a division of labor. This division lets developers fulfill a task appropriate to their profession and lets different tasks proceed in parallel.

Finally, the application architecture proposed in this paper is based improves the development efficiency of e-Learning system. The architecture is simple, easy to use and across languages, frameworks and runtime platforms, it not only can avoid the repeated writing the similar logic of each domain module, also is conducive to the robustness, maintainability of the system, and flexibility to the changes of whole business needs. It will have an important guiding significance for the application and development of the e-Learning system in the network education.

Acknowledgment. This work was supported by the Science and Technology Project of Education Department of Liaoning Province, China (①). Research of e-Learning System of small and medium-size Enterprises Based on SaaS, No.L2013417. ②. Research on strategy for balanced development of elementary education, No.JG12CB215).

References

1. Goel, N., Maitrey, S., Kanauzuya, S.: Web Technologies (Web2.0). *International Journal of Computer Applications* 1(6), 11–12 (2010)
2. Song, B., Li, M.Y.: An e-Learning System Based on GWT and Berkeley DB. In: Tan, Y., Shi, Y., Ji, Z. (eds.) *ICSI 2012, Part II. LNCS*, vol. 7332, pp. 26–32. Springer, Heidelberg (2012)
3. Song, B., Liu, J.: Implementation of J2EE Data Persistence Tier with TopLink. *Microelectronics & Computer* 23(8), 132–135 (2006)
4. Song, B., Zhao, J.: Research on Network Teaching System Based on Open Source Framework. In: *IEEE Ninth International Conference on Hybrid Intelligent Systems*, vol. 1, pp. 28–32. IEEE, Shenyang (2009)
5. Fang, W., Sun, Y.: Research and Application of J2EE's Data Persistence Layer. *Computer Technology and Development* 17(2), 68–91 (2007)
6. Song, B., Zhang, Y.: Implementation on Network Teaching System Based on Java EE Architecture. In: *IEEE Second International Conference on Information Technology and Computer Science*, Kiev, vol. 1, pp. 354–357 (2010)
7. IBM Corporation, *EGL Programmer's Guide Version7 Release 00*, USA (2007)
8. Russell, M.A.: *DOJO: The Definitive Guide*. O'Reilly Media, USA (2009)
9. Margolis, B., Danny, A.: *Enterprise Web 2.0 with EGL*, pp. 83–95, 131–141, 225–237. MC Press, USA (2009)

Adaptive Pulse Design and Spectrum Handoff Technology Based on Cognition*

Bing Zhao**, Erfu Wang, Jiaqi Zhen, and Qun Ding

Electronic Engineering, Heilongjiang University
Harbin, Heilongjiang, China
zb0624@163.com

Abstract. Cognitive ultra wideband wireless communication system is a new type of intelligent communication system, which provides an effective method to relieve the shortage of the spectrum resource. In order to avoid interference with existing communication systems, this new one has a high demand for pulse design and spectrum handoff. This paper presents a flexible multi-band adaptive pulse design method and a proactive spectrum handoff mechanism which applies to the pulse above, the aim of that is expect to reduce the spectrum loss during the spectrum handoff and improve spectrum utilization. Experimental results show that on the condition of one cognitive user channel only hold a licensed user, the pulse spectrum utilization can reach above 80%, the system spectrum utilization is not affected by the arrival rate of licensed users.

Keywords: cognitive ultra wideband, adaptive pulse design, spectrum utilization, proactive spectrum handoff, part avoidance.

1 Introduction

Spectrum resources demand become increasing with the rapid development of wireless communications technology, the non-effective use of the original licensed spectrum led to the spectrum resources tension. Opening unused band and conditionally limiting the unlicensed user intervention is an effective way to solve the lack of spectrum resources and improve the spectrum utilization. Cognitive ultra wide-band(CUWB) is an intelligent Wireless communication technology, which combine the cognitive radio(CR) and the ultra wide-band technology. CR user can occupied idle spectrum resources temporarily by sensing the spectrum environment, adaptive changes in operational parameters, the specific way is random insert. While licensed users with a higher priority appears or the channel state can not be satisfied with business requirements, CR users must exit the current spectrum, looking for new spectrum hole and re-establish the link in order to maintain ongoing communication, this process is called spectrum

* This work is supported by Heilongjiang Provincial Education Department Science and Technology Research Project (NO.12531492). Many thanks to the anonymous reviewers, whose insightful comments made this a better paper.

** Corresponding author.

handoff process, caused by the movement of spectrum. During the handoff, if there are no idle spectrum holes or the spectrum characteristics can not satisfied QoS of CR user services, the communication will be interrupted, which should try to avoid for cognitive system. The current study is aimed at narrowband communication in primarily cognitive radio environment; research on for ultra-wideband signal spectrum handoff strategy in CUWB system is still very little. This essay presents a flexible design methods and the corresponding spectrum handoff mechanism for wideband cognitive users[1,2,3].

2 CUWB Adaptive Pulse Design

The traditional UWB signal mainly in the form of narrow pulses can be used in many different waveforms, such as Gaussian waveform, raised cosine waveform, etc. These are mostly based on the waveform from the time domain to the frequency domain design ideas. B.Parr, who first proposed the frequency domain to the time domain design ideas, that, evaluating the time-domain expression of pulse which based on the frequency characteristics that the ultra wide-band must satisfies FCC emission mask[4,5].

In order to satisfy the radiation mask requirements of signal power spectrum density and reach to high-speed transmission and low inter-symbol interference, CUWB system requires a pulse signal both band-limited and time-limited. Prolate spheroidal wave function (PSWF) is the complete orthogonal basis in band limited space $[-\Omega, \Omega]$ and time limited space $[-\frac{T}{2}, \frac{T}{2}]$ [6,7,8,9]. It satisfies the following integral equations:

$$\int_{-\frac{T}{2}}^{\frac{T}{2}} \varphi(x) \frac{\sin \Omega(t-x)}{\pi(t-x)} dx = \lambda \varphi(t) \tag{1}$$

$$\int_{-\frac{T}{2}}^{\frac{T}{2}} \varphi_i(x) \varphi_j(x) dt = \begin{cases} \lambda, & i = j \\ 0, & i \neq j \end{cases} \tag{2}$$

$\varphi(x)$ is prolate spheroidal wave function, λ is the corresponding eigenvalue, $1 > \lambda_0 > \lambda_1 > \dots > \lambda_i > \dots, \lambda_n$ is the energy concentration of output pulse. They all determined by the time-bandwidth product $C = \frac{T\Omega}{2}$.

$$\lambda = \frac{\int_{-\frac{T}{2}}^{\frac{T}{2}} |\varphi_n(t)|^2 dt}{\int_{-\infty}^{\infty} |\varphi(t)|^2 dt} \tag{3}$$

Represents the equation (1) into discrete matrix form:

$$\mathbf{H}\varphi = \lambda\varphi \tag{4}$$

In above questions: \mathbf{H} is hermite matrix, φ is different eigenvalue λ corresponding to eigenvector groups.

CUWB adaptive pulse design ideas are divided the available frequency band into several sub-bands, using PSWF produce the corresponding sub-band pulse

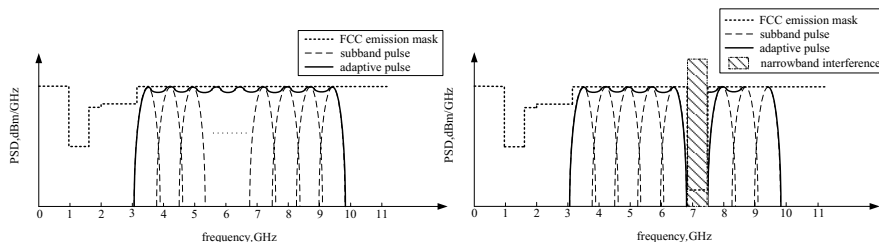


Fig. 1. Multi-band pulse design principle

waveform and sub-band pulse will be weighted sum in the time domain, devising a multiband adaptive pulse with different central frequencies. Multi-band spectrum pulse system design flexibility, with a strong coexistence and rules of adaptability, when there is a narrowband interference licensed user or existing interruptions, in order to avoid mutual interference, making the sub pulse add up to zero in the interference bands by prohibiting the use of some sub bands or adjusting parameters, thus avoiding interference, multi-band pulse system shown in figure 1.

Design steps as follows:

(1) To build an adaptive radiation mask. Before design the adaptive pulse, first sensing the radio spectrum environment, selecting the band which satisfies communication requirements, constructing adaptive radiation mask dynamically, which required by FCC radiation mask. CUWB system with has better flexibility than fixed radiation mask for UWB systems.

(2) Set the pulse related parameters. CR user determines the position of the divided sub-band $[f_{iL}, f_{iH}]$ and the number of sub bands M based on the spectrum bandwidth of the pulse B , the pulse duration T_m , the time bandwidth product C . In order to make better use of spectrum resources, improving the spectrum utilization, while dividing each sub-band, the band between each sub-band is overlapping, namely $f_{iH} > f_{(i+1)H}$, the definition of the band overlap rate v :

$$v = \frac{(f_{(i-1)H} + f_{iH}) - (f_{iL} + f_{(i+1)L})}{2 \times (f_{iH} - f_{iL})} \times 100\% \tag{5}$$

Assuming each sub-band has the same width B_0 , through the formula $B_0 = \frac{C}{T_m}$, then calculate the width of the sub-band, determine the number of sub-band M by the following formula:

$$B = B_0[M - v(M - 1)] \tag{6}$$

Sub-band pulses are selected from the top m PSWF with larger concentration energy, you can design m^M multi-band pulse in theory.

(3) Sub-band division and build sub-band radiation mask. The band is divided into M sub-bands, and build sub-band radiation mask based on the spectrum

characteristics of each sub-band to control the pulse power, making the power spectrum density of the multi-band adaptive pulse does not exceed the specified range $S_{CUWB}(f)$.

(4) Design sub-band pulse $p_i(t)$. Let the i -th sub-band spectrum range be $[f_{iL}, f_{iH}]$, in all the waveforms that satisfy the conditions, selecting the PSWF φ_k with energy concentration λ_k as a sub-band pulse $p_i(t)$.

$$p_i(t) = \varphi_{ik}(t), \quad i = 1, 2, \dots, M, \quad k = 0, 1, 2, \dots, m - 1 \quad (7)$$

(5) Making weighted sum on sub-band pulse for each in the time domain, getting multi-band adaptive pulse $p(t)$ after weighted sum on the generated sub-band pulse in the time domain.

$$p(t) = \sum_{i=1}^M b_i p_i(t - \tau_i) \quad (8)$$

In above questions: b_i is weighting coefficient of the corresponding sub-band; τ_i is phase factor, making $\tau_i = 0$.

Focus that the sub-band pulse phase is cyclical, the number of cycles is $p = \frac{T_m B_0}{2}$. The principle of selecting sub-band pulse is: under the condition of the energy with the greatest possible, selecting the sub-band pulse phase with the same or similar phase in overlapped spectrum to add up. In order to ensure the sub-band pulses with same phase can be founded, on the condition that each sub-band has the same width, overlap rate v should satisfy the following formula :

$$v = \frac{i}{4p} \times 100\%, \quad i = 1, 2, \dots, 2p \quad (9)$$

Figure 2(a) is an adaptive pulse power spectrum density with pulse duration $T_m = 2ns$, available frequency band $f \in [1.5GHz, 5GHz]$, the number of sub-band 8, the pulse frequency band overlapping 50%. Figure 2(b) is an adaptive pulse power spectrum density while narrow-band interference occurs, when the pulse in the frequency band $[2.7GHz, 3.3GHz]$ get a groove within a depth of $40dB$, these grooves can make CUWB wireless communication system avoid interfering the narrowband systems on this band, so as to be coexistence.

Calculate the pulse spectrum utilization is one way to measure the pulse of the merits of CUWB, spectrum utilization is defined as:

$$\eta = \frac{\int_B S_p(f)df}{\int_B S_{CUWB}(f)df} \times 100\% = \frac{\int_B \frac{1}{T_s} |P(f)|^2 df}{\int_B S_{CUWB}(f)df} \times 100\% \quad (10)$$

In above questions: T_s is the pulse repetition period.

Spectrum utilization of adaptive pulse is influenced by pulse duration T_m , time-bandwidth product C , the number of sub-bands M and band overlap rate v . And in the case of T_m and C is determined, the spectrum utilization increases with the increasing of the band overlap ratio v , but the increase of v will cause an increase in the number of sub bands M , result in the increasing of calculation and system complexity.

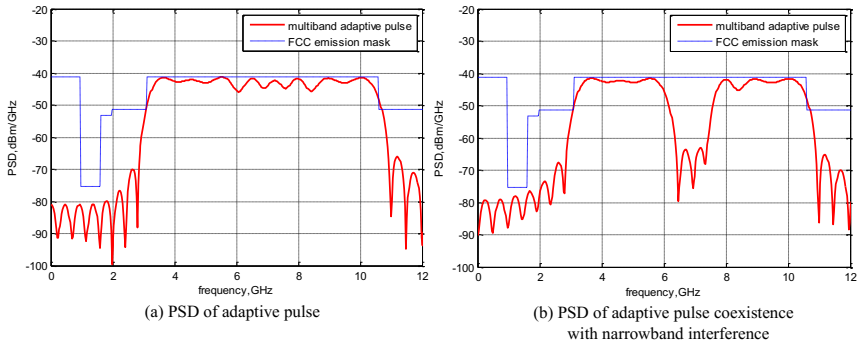


Fig. 2. PSD of adaptive pulse

Table 1. Comparison of the pulse spectrum utilization

Pulse waveform	spectrum utilization %
Modified Gaussian 5/7 order derivative pulse	50.90
33 combinations of Gaussian pulse	92.16
Optimized Rayleigh 4/6 order derivative pulse	51.94
Hermite 1 order correction derivative pulse	46.23
Morlet orthogonal wavelet pulse	30.70
19 combinations Chirp compression pulse	83.33
Avoid co-channel interference 17 combinations Chirp compression pulse	73.30
Multi-band adaptive pulse	85.83

Table 1 is the contrast between adaptive pulse and traditional UWB pulse waveform spectrum utilization. As can be seen from the table, the adaptive pulse spectrum utilization can be guaranteed more than 80%, even in the same time of avoiding the licensed user.

Due to the very wide spectrum of ultra wide-band pulse, interfere is inevitably with existing narrowband wireless systems. When a small amount of narrowband interference occurs in the available bandwidth, we only need to cancel the sub-band pulse in the interfering bands instead of changing the communication bands. So as to realize revising seamless transmitted waveform to adapt with the wireless environment. Adaptive pulse can still ensure over 80% spectrum utilization while escaping from the licensed users successfully.

3 Spectrum Handoff

When the CR users spectrum occurs movement and handoff in the communication process, the communication links will be interrupted temporarily, after CR users find new spectrum holes, the transmitter and the receiver must re-establish

the link. Transmitting terminal conducts the spectrum handoff and informs the receiver by handshake protocol, this process will bring new communication overhead, that is handoff delay[10]. For the node in a distributed network, designing a efficient and fast handshake mechanism between transceivers is particularly important.

3.1 Spectrum Pooling Strategy

CR users should do regular spectrum detection and record the testing results in accordance with the strength of business, whether in communicate state or not. CR users select L_{max} spectrum holes from N available spectrum holes that find in each round to build their own spectrum pool. This process can ensure CR users don't need to re-test the spectrum when a licensed user with a higher priority occurs, they can continue to communicate as long as selecting the appropriate spectrum holes from the pool they recorded.

For the spectrum resources in the pool, CR users and licensed users have different usage rules: the business from CR users can only use the resources that in the spectrum pool, but the business from licensed users is unlimited, not only can use the resources in spectrum pool but also using it beyond the spectrum pool. But no matter what kind of business occupied the spectrum holds, the resources that release in the end of the business will not belong to the pool until the next spectrum detection, that is to say, before next spectrum detection begin, CR users can only know when spectrum holds will be occupied, but do not know how long it will be occupied. Therefore, as time goes by, the resources in the pool will reduce, and when spectrum holds reduce to threshold L_{min} , CR users begin to start a new round of spectrum detection.

The reduce rate γ in the pools is mainly affected by the licensed users access rate γ_a and CR users access rate γ_b , at any time, when the number of pool's users is k , the total reduce rate of the spectrum pools γ is:

$$\gamma = \lceil \frac{B_c}{B_l} \rceil \gamma_b + \frac{L_{max} - k}{N} \gamma_a, B_l < B_c \quad (11)$$

In above questions: B_l refers to licensed users signal bandwidth, B_c refers to CR users signal bandwidth, $\lceil \cdot \rceil$ refers to rounding up operation.

The bigger the reduce rate γ , the more business volume in this area, and the shorter interval between two spectrum detection. If the result of spectrum detection is the numbers of available spectrum holds N less than the set L_{max} , all the resources will belong to the spectrum pool, that is , $L_{max} = N$; If N less than the set L_{min} , the system cannot set up the spectrum pools.

3.2 Part Avoidance Mechanism

In CUWB system, combined with the requirement of real-time, introducing buffer zone that based on the spectrum pool strategy, and proposed an available part avoidance mechanism for broadband cognitive users. Suppose the number of

channels in spectrum pools is S , a CR user occupied the number of channel is n , and one CR user's channel at most can be shared with one narrowband licensed user, two licensed users can not occupied adjacent channels at the same time, and the spacing of the channel occupied is large, ensure that each CR user channel only has one licensed user in probability, the capacity of buffer and spectrum pool is the same, following the first-in first-out service rules, the longest queuing time is τ .

When licensed users appear, CR users which in the same channel will first estimate the signal broadband of that, and calculating the bandwidth which needed to avoid, determining whether they can coexist with licensed users or not, the ratio of the signal bandwidth is:

$$\frac{B_l}{B_c} = \frac{1}{\rho} \tag{12}$$

In above questions: B_c refers to the bandwidth of CR users, B_l refers to the bandwidth of licensed users, that is, the bandwidth that CR users need to avoid.

Assuming that the service request time of licensed users and CR users obey the negative exponential distribution, the average duration of each business time is $\frac{1}{\mu_a}$ and $\frac{1}{\mu_b}$. r is the threshold that CR users set based on the signal power and transmission distance, if $\rho \geq r$, CR users do not need to handoff channels but only need to remise partial bands that licensed users needed, adjusting transmission parameters at same time and continuing to communicate by the remaining bands, otherwise, CR users are forcibly occupied the channels, exiting from the communication immediately, returning all the channels and queuing in buffer to wait for other idle channels. If the CR users who queued in the buffer still not has access services during the maximum waiting time τ , then CR users will be forced interrupted. In the current spectrum of CUMB open range, the licensed users are satisfied with formula (12), at that time, the spectrum pool can hold at least $\frac{S}{n}$ licensed users and $\frac{S}{n}$ CR users, the buffer can hold $\frac{S}{n}$ CR users to queue. Under certain conditions, CR users can coexist with licensed users. Transition probability of the neighboring state in partial avoidance mechanism is:

$$R_{(i,j) \rightarrow (i,j+1)} = \lambda_a, \quad 0 \leq i \leq \frac{S}{n}, 0 \leq j \leq \frac{S}{n} - 1 \tag{13}$$

$$R_{(i,j) \rightarrow (i,j-1)} = j\mu_a, \quad 0 \leq i \leq \frac{S}{n}, 0 \leq j \leq \frac{S}{n} \tag{14}$$

$$R_{(i,j) \rightarrow (i+1,j)} = \begin{cases} \lambda_b, & 0 \leq i \leq \frac{S}{n} - 1, 0 \leq j \leq \frac{S}{n} \\ 0, & \frac{S}{n} \leq i \leq \frac{2S}{n}, 0 \leq j \leq \frac{S}{n} \end{cases} \tag{15}$$

$$R_{(i,j) \rightarrow (i-1,j)} = \begin{cases} i\mu_b, & 1 \leq i \leq \frac{S}{n}, 0 \leq j \leq \frac{S}{n} \\ \frac{S}{n}\mu_b + \frac{(i-\frac{S}{n})}{\tau}, & \frac{S}{n} \leq i \leq \frac{2S}{n}, 0 \leq j \leq \frac{S}{n} \end{cases} \tag{16}$$

Figure 3 is a flow diagram of part avoidance mechanism.

Simulated on the spectrum utilization for any CR users within the user bandwidth range, and set the signal bandwidth of licensed users is 20 100MHz, with

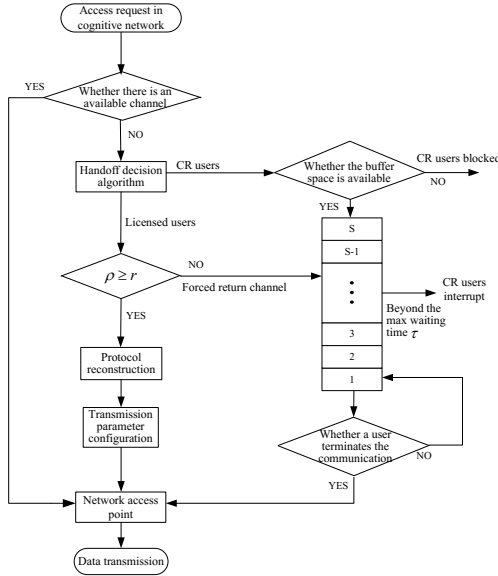


Fig. 3. Flow diagram of part avoidance mechanism

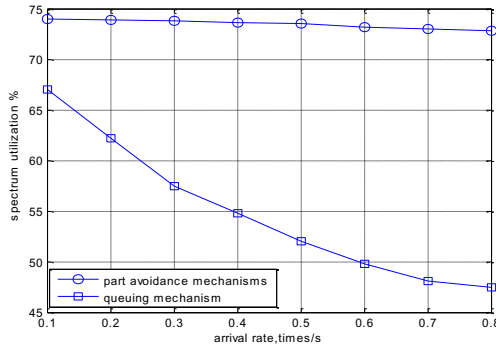


Fig. 4. Changes of spectrum utilization with license user arrival rate under different handoff mechanism

an average duration $\frac{1}{\mu_a} = 400ms$, and CR user’s signal bandwidth is $1GHz$, using adaptive pulse, $T_m = 2ns$, the data transmission rate is $22Mbit/s$, the average duration is $\frac{1}{\mu_b} = 40ms$, the arrival rate is $20time/s$, the simulation time is $5s$.

Figure 4 is the change situation of the system average spectrum utilization with the licensed user arrival rate on the condition of using part avoidance mechanisms and traditional request queuing mechanism. As the figure shows: as the

arrival rate of licensed users increasing, the spectrum handoff of CUWB system which using request queuing mechanism increase, the spectrum utilization of licensed user is very low, so the average spectrum utilization of the system reduce, system cannot make full use of spectrum resources; the spectrum utilization of CUMB which using part avoidance mechanism is not affected by the arrival rate of licensed users basically, maintaining a high spectrum utilization in simulation time and making full use of spectrum resources.

4 Conclusion

This paper research on the wide band adaptive pulse design method based on cognition and spectrum handoff mechanism applying to wide band cognitive users, the simulation results show that when an licensed user and cognitive user bandwidth ratio under the allowable threshold, they can coexist by adjusting sub-band pulse. Reducing the spectrum handoff frequency, to achieve the purpose of improving the spectrum utilization, at the same time, keeping the business quality of cognitive user not decline, providing basic research results for improving spectrum utilization and for solving the new challenges faced by cognitive wireless networks, the future research focuses on the cognitive wireless network technology and the integration of future ubiquitous wireless networks.

References

1. Mitola, J., Maguire, G.Q.: Cognitive Radio: Making Software Radios More Personal. *Personal Communications* 6, 13–18 (1999)
2. Chen, X.B., Chen, H., Chen, Y.: Spectrum Sensing for Cognitive Ultra-wideband Based on Fractal Dimensions. In: 2011 Fourth International Workshop on Chaos-Fractals Theories and Applications, pp. 363–367. IEEE Press, Hangzhou (2011)
3. Masri, A.M., Chiasserini, C.F., Casetti, C., Perotti, A.: Common Control Channel Allocation in Cognitive Radio Networks through UWB Communication. *J. Commun. Netw.-S Kor.* 14, 710–718 (2012)
4. Li, B., Zhou, Z., Zou, W.X.: A Novel Spectrum Adaptive UWB Pulse: Application in Cognitive Radio. In: 2009 IEEE 70th Vehicular Technology Conference Fall, pp. 1–5. IEEE Press, Anchorage (2009)
5. Yang, J., Jian, X.J., Hou, Z.F.: Interference Suppression by Adaptive Pulses for Cognitive UWB. In: 2008 9th International Conference on Signal Processing, pp. 1780–1783. IEEE Press, Beijing (2008)
6. Slepian, D.: Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty. *Bell System Technical Journal* 40, 43–64 (1978)
7. Walter, G.G., Shen, X.P.: Wavelets Based on Prolate Spheroidal Wave Functions. *J. Fourier Anal. Appl.* 10, 1–26 (2004)
8. Zhang, H.G., Kohno, R.: SSA Realization in UWB Multiple Access Systems Based on Prolate Spheroidal Wave Functions. In: 2004 IEEE Wireless Communications and Networking Conference, pp. 1794–1799. IEEE Press, Atlanta (2004)

9. Karoui, A., Moumni, T.: New Efficient Methods of Computing the Prolate Spheroidal Wave Functions and Their Corresponding Eigenvalues. *Appl. Comput. Harmon. A.* 24, 269–289 (2008)
10. Wang, P., Xiao, L.M., Zhou, S.D., Wang, T.: Optimization of Detection Time for Channel Efficiency in Cognitive Radio Systems. In: *IEEE Wireless Communications and Networking Conference*, pp. 111–115. IEEE Press, Hong Kong (2007)

Technology Research of the Configured Component ERP System Based on XML

Jialin Ma¹ and Yi Hou²

¹ Shenyang Normal University, 110034, Shenyang, China

² Shenyang Radio and TV University, 110003, Shenyang, China
jialinma@126.com

Abstract. Development of component technology to the development of ERP systems and opportunities. This paper proposes an XML-based ERP system components can be configured for rapid b. Through software reuse, component-based XML configurable ERP system to achieve rapid development, away from the traditional code of programming complex and cumbersome process, completely by calling the XML components to achieve, not only can quickly generate interface design, and business process design is also XML component settings, even the function of enterprise systems updates, as long as you can by modifying the XML components, thereby improving software productivity, improve system performance, reduce development costs and maintenance costs.

Keywords: ERP systems, rapid development, XML, configurable components, software reuse.

1 Introduction

ERP is a supply chain management-oriented information system, which uses computer technology, network communication technology, the implementation of customer-centric strategy elite, considering the manufacturers, suppliers, distributors and other customers in all aspects of a combination of factors to achieve rational allocation of corporate resources. In recent years, most enterprises have started the application and implementation of ERP systems, ERP systems with the further development of information technology and modern management thinking, ERP system will continue to improve.

2 Key Technology Components Based on XML Configurable ERP Systems

2.1 Software Reuse

Software reuse (Software Reuse) is a variety of existing knowledge about the software used to create new software to reduce the cost of software development and maintenance. Software reuse is an important technology to improve software

productivity and quality. The main idea is: a software as a function of a number of different components of the composition, each component is equivalent to a universal tool can be completed the same function, so if this type of assembly in accordance with the functional requirements of rational organization linking is complete preparation of a specific software, these components are not only used in this software, but can be reused for other related software.

2.2 Component Technology

Component technology is the use of some of the programming method, some are not easy for users to directly manipulate the details of the package, while achieving a variety of business logic. In the actual development process, each component will provide some standard application interface, the user can adjust its parameters, while the needs of software developers will be functional components together organically, quickly developed an application that meets the actual needs . Component-based development model multiplexed (Fig 1).

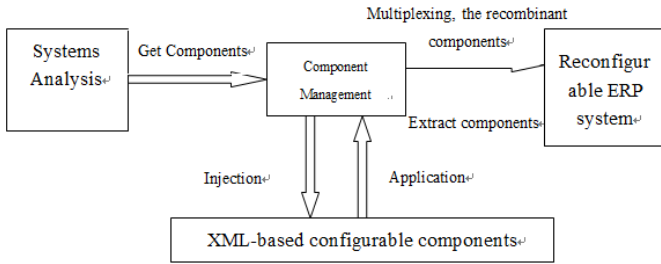


Fig. 1. Component-based software development

2.3 XML Technology

XML (eXtensible Markup Language) is a standard for data conversion and represented World Wide Web Consortium (W3C) 1998 was released, and is a universal markup language a subset of SGML, a markup language as it has a simple, flexible, widely used and powerful characteristics. More extensive use of XML, its power comes from its data independence. For the application of XML technology mainly involves two aspects: XML parsing and XML data processing.

3 XML Technology Components Can Be Configured in the ERP System Development and Implementation

3.1 XML Data Processing Techniques

To achieve internal and external data exchange ERP system, you need to have a data transfer to pay the "software" that can exchange data between databases and XML, it

is the task of ERP systems: extract data from the current database and generate XML documents; and the received XML document stored into a relational database; standard XML document as a transmission medium, which can interact with the data between different databases.

3.1.1 Extract XML Standard for Data Exchange Documents

XML document has "to self-describing , " " infinite nesting , " " tree ," and so , in a sense , an XML document is a one of a database or table . Standard XML document as a transmission medium for data exchange between databases could not be better , the standard XML document needs to dependencies between the mark and the mark is defined , so that users clearly labeled dependencies between meaning and markers , and need to define a unified data format , only the use of a unified data format , in order to achieve the free exchange of data between internal systems and processes as well as different databases.

3.1.2 Mutual Conversion between Database and XML Script

If you want to achieve data compatibility between the different components of the database and the database, you must first convert between XML documents and databases of data between XML schemas and database schema mapping with each other, where we use the table-based mapping to achieve two mutual mapping problem between modes. Here can be achieved through a conversion script specific conversion between relational databases and XML, a relational database for specific , according to the form DTD or XML Schema to define the conversion script . After defining the data conversion scripts , data conversion components use these scripts to convert between relational databases and XML documents conversion data .

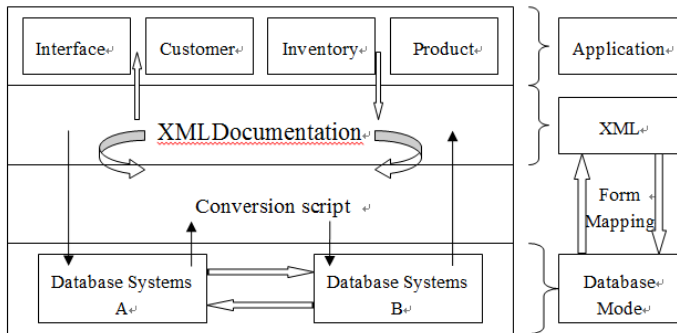


Fig. 2. Illustrates the data mode conversion

3.2 XML Components Can be Configured to Achieve

3.2.1 Component Manager Software Bus

Component Manager to complete what the main functions: (1) check the identity and security of the individual components ; (2) deploy , activate , upgrade, and suspend and delete the specified data processing components ; (3) legitimate software

components are organized in a bus (XML) text ; (4) when the system starts again when the components of the bus automatically checked and activation ; (5) only in the bus components can be used.

3.2.2 Database Access Components

This section includes additional single database table , delete , change, and other atomic components, complex transaction processing components and assembled using atomic components made of composite components . Because the database table is specific for each lot and is in the system , so to build a common atomic database access components. For component design involves two main aspects :

(1) database connection pool : access component may also be called by multiple applications section , to access different databases and forms may also do different operations. Each operation should be carried out in order to avoid connecting to the database , access and disconnect , you need to add a system management database connection pool based on XML configuration files, to unify the management and use of the database to connect to database components and systems to improve access efficiency.

(2) business data : This design uses XML database provides new data types xml Type to store business data templates for each business and business data instances and use SQL and XQuery technology to achieve access to business data and internal information processing . In this way , the DBMS 's own mechanisms to ensure the integrity of data accessed , the above also applies to the design of the database access components.

3.2.3 Data Processing Combination of Components

Sometimes, some of the data processing operations are very complex and require access to different databases , different data forms , you need to combine this type of atomic components and other components of these transactions can be achieved . Under normal circumstances, such a component is acceptable to meet the task (can be broken down into a series of simple tasks can be completed by the atomic components) , and in a certain order to perform these simple tasks . Combination of components from the XML-based bus management portfolio atomic components (configurable) , atomic components (referenced assembly) , task scheduling component (to decide when to call the appropriate components) , the interface components (scheduling algorithm can be used) .

3.2.4 XML Parsing and Generation Components

Data layers are accepted XML-based description of a unified data access tasks, each database access components are required for the task of parsing XML and database components can recognize and generate SQL-based database access language statement processing. Since all the data layer components are returned XML-based description of the task , you need to generate XML components to the results returned by the XML package.

4 Conclusion

In this paper, XML-based ERP system can be configured components for rapid development of enterprise information technology undoubtedly has greatly promoted the development of, so that enterprises in the current market demand for fast and flexible in a timely grasp opportunities. Although you can configure the XML-based component is currently still not very mature, positive so it was with great development space. XML can be configured in the current assembly methods have been exposed edge applications, such software obviously has good flexibility, more suitable for the current software vendors and enterprises.

References

1. Maitin, D.: XML Advanced Programming, pp. 55–59. Mechanical Industry Press, M. Beijing (2001)
2. Jian, W.: XML data processing techniques used in ERP. J. Computer Knowledge and Technology (2010)
3. Huang, Z.M., Xue, H., Gui, L.J.: Reconfigurable ERP software development technology (2006)
4. Attentive, high-mao court. XML reconfigurable manufacturing execution system component management application (2006)
5. XML database technology, <http://www.ibm.com/>
6. XML Schema Infoset Model, <http://www.xml.org.cn/index.html>

Discriminative Feature Learning for Action Recognition Using a Stacked Denoising Autoencoder

Ruoxin Sang, Peiquan Jin, and Shouhong Wan

School of Computer Science and Technology,
Key Laboratory of Electromagnetic Space Information, Chinese Academy of Sciences,
University of Science and Technology of China, Hefei, China
srx2007@mail.ustc.edu.cn, {jpeq,wansh}@ustc.edu.cn

Abstract. In this paper, we propose a novel method to recognize human actions based on the depth information acquired by depth-based cameras. Representations of depth maps are learned and reconstructed using a stacked denoising autoencoder. By adding the category constraint, the learned features are more discriminative and able to capture the small but significant differences between actions. Greedy layer-wise training strategy is used to train the deep neural network. Then we use temporal pyramid matching on the feature representation to generate temporal representation. Finally a linear SVM is trained to classify each sequence into actions. Our method is evaluated on MSR Action3D dataset and show superiority over other popular methods. Experimental results also indicate the great power of our model to restore highly noisy input data.

Keywords: Action Recognition, Feature Learning, Stacked Denoising Autoencoders.

1 Introduction

Human action recognition has been an active field of research in computer vision. The goal of action recognition is to recognize people's behavior from videos in a given scenario automatically. It has many potential applications including content-based video search, human computer interaction, video surveillance, sports video [1, 2]. Most of these applications require high level understanding of spatial and temporal information from videos that are usually composed of multiple simple actions of persons.

Inferring high-level knowledge from a color video especially in a complex and unconstrained scene is very difficult and costly. However, the recent availability of depth cameras such as Kinect [3] has tremendously improved the abilities to understand human activities. Depth maps have several advantages over traditional intensity sensors. First, depth sensors can obtain the holistic 3D structure of the human body, which is invariant to color and texture. Second, color and texture methods perform worse in the dim lighter and the shadows may bring

ambiguity. But the depth cameras can work in total darkness. Third, depth sensors greatly simplify the process of foreground extraction, removing plenty of noise and disturbance in the background [4, 5].

Furthermore, the 3D skeleton joint positions can be estimated from the depth map accurately following the work of Shotton *et al.* [3]. The extracted skeleton joints have strong representation power, which is more discerning and compact than depth or color sequences. Although with these benefits, depth-based action recognition using joint features is still not an easy task [6]. Some of the estimated joints are not reliable when the human body is partly in view. The overlapping of human parts in some interactive actions can lead to the missing of some joint as well. Due to the noisy joint positions, extracting robust features from skeleton information is necessary.

Motivated by the satisfactory performance of previous work on exploring relative 3D joint features [2, 7, 8], we propose a novel method to learn robust and discriminative features from joint 3D features to recognize human actions. We build a deep neural network and employ denoising autoencoders, which has proved their strong abilities to reconstruct and denoise data, as the basic unit of our architecture. In order to seize very subtle spatio-temporal details between similar actions, we add the category constraint on denoising autoencoders to fuse intra-and inter-class information into features. We stack the denoising autoencoders with category constraint and greedy layer-wise training strategy is used to train the model. Then we use temporal pyramid matching on the feature representation to generate temporal representation. Finally a linear SVM is trained to classify each sequence into actions. Experiments show that this algorithm achieves superior results on a benchmark dataset.

The main contributions of this paper are three-fold. First, a new discriminative feature learning algorithm is proposed to recognize depth-based videos. Second, a novel category constraint is added into denoising autoencoders to preserve intra-and inter class information. Third, our extensive experiments show that our model has a strong capacity to reconstruct and denoise corrupted data.

The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 describes the entire flow of our methodology to recognize actions. Section 4 discusses the experimental results. Section 8 concludes the paper.

2 Related Work

Recently, low-level hand-crafted features have been designed to recognize human actions. Spatio-temporal salient points like STIP [9] or some local features, like Cuboids [10] and HOG\HOF [11] have been widely used. However, directly employ these original methods for color sequences on depth data is infeasible. Therefore, recent methods for action recognition in depth sequences explore alternative features particularly for depth-based videos. Li *et al.* [12] projected the depth map into three orthogonal planes and sampled representative 3D points to obtain a bag of 3D points. An action graph was deployed to model the dynamics

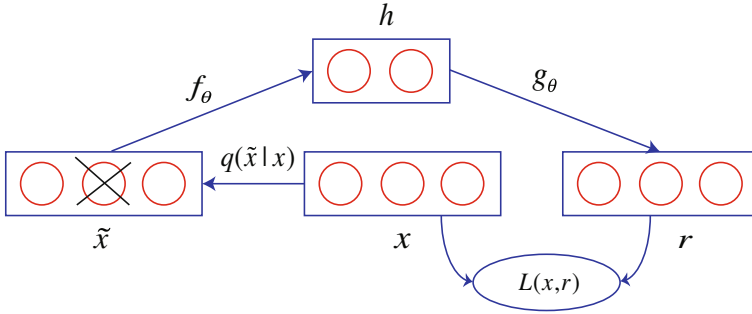


Fig. 1. The architecture of the denoising autoencoder. The input data x is stochastic corrupted into \tilde{x} by mapping function $q(\tilde{x}|x)$. The autoencoder then maps \tilde{x} to h and maps back h to r , the reconstruction result. $L(x,r)$ is the reconstruction error measurement function.

of the salient postures. Lu *et al.* [4] extracted spatio-temporal interest points from depth videos and built a cuboid similarity feature. Similarly, in [5], Omar and Zicheng quantized the 4D space and represented the possible directions for the 4D normal in order to build a histogram in the 4D space.

As mentioned before, skeletal information has strong representation power. Lu *et al.* [7] computed histograms of 3D joint locations, reprojected the extracted features using LDA [13], and clustered them into visual words. The temporal evolutions of these words were modeled by HMMs [14]. Jiang *et al.* [2] combined skeleton and depth information to obtain Local Occupancy Patterns (LOP) at each joint and built a Fourier Temporal Pyramid, an actionlet ensemble was learned to represent the actions. Jiajia [6] proposed a dictionary learning algorithm adding the group sparsity and geometry constraints, obtain an overcomplete set of the input skeletal features. The Temporal Pyramid Matching was used for keeping the temporal information.

Deep Learning [15–18] is a set of algorithms that attempt to learn a hierarchy of features by building high-level features from low-level ones. Some models such as CNN [18], DBN [16] and Autoencoders [15] have achieved surprising results in areas like computer vision, natural language processing and speech recognition. One reason for the success of deep learning methods is that they usually learn to capture the posterior distribution of the underlying explanatory factors for the data [19]. Therefore, rather than elaborately designing the hand-crafted features as in [5], we choose to learn high level features from data. The experimental results further prove the feasibility and validity of deep learning methods.

3 Proposed Method

In this section, we will first describe the basic Denoising Autoencoders. Next, we will extend the model by adding the category constraint, to make the learned features more discriminative and obtain better accuracies for recognizing actions.

Then we introduce the stacking techniques to build a deep architecture. Finally, we employ temporal pyramid matching to generate the temporal representation and do classification.

3.1 Denoising Autoencoders

Autoencoders were proposed by Hinton [15] to recognize handwritten digits, which achieved the state of the art at that time. An autoencoder is a special kind of neural networks whose target values are equal to the input ones. A single-layer Autoencoder comprises two parts: **encoder** and **decoder**.

Encoder: The transformation function maps an input vector x into a hidden layer feature vector h . Its typical form is a non-linearity function. For each example $x^{(i)}$ from a data set $\{x^{(1)}, x^{(2)}, \dots, x^{(n)}\}$, we define:

$$f_{\theta}(x^{(i)}) = s(Wx^{(i)} + b) \quad (1)$$

Decoder: The parameterized function maps the hidden layer feature vector h back to the input space, producing a reconstruction vector:

$$g_{\theta}(h^{(i)}) = s(W'h^{(i)} + c) \quad (2)$$

The set of parameters of this model is $\theta = \{W, W', b, c\}$, where W and W' are the encoder and decoder weight matrices and b and c are the encoder and decoder bias vectors. It is worth mentioning the input vector $x^{(i)}$ and the reconstruction vector $r^{(i)}$ have the same dimension d_x , the hidden layer $h^{(i)}$ has the dimension d_h , thus the size of W is the same as the size of transpose of W' , which is $d_h \times d_x$.

The basic autoencoders aim to minimize the reconstruction error of all samples:

$$L_{AE}(\theta) = \sum_i L(x^{(i)}, g_{\theta}(f_{\theta}(x^{(i)}))) \quad (3)$$

In practice, the choice of function s is usually a sigmoid function $s(x) = \frac{1}{1+e^{-x}}$ or a tanh function $s(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ and the loss function L is usually a square loss function $L(x, r) = \|x - r\|^2$.

Vincent [20] proposed Stacked Denoising Autoencoders (SDA), exploring a strategy to denoise corrupted version of input data. The input x is first corrupted into \tilde{x} using stochastic mapping $\tilde{x} \sim q(\tilde{x}|x)$. This is like randomly selecting some nodes of the input and blinding them, that is, every node in the input layer has a possibility q to be switched to zero. The stochastic corrupted data is regarded as the input of next layer, see Fig. 1. This yields the following objective function:

$$L_{DAE}(\theta) = \sum_i \mathbb{E}_{q(\tilde{x}|x^{(i)})} \left[L(x^{(i)}, g_{\theta}(f_{\theta}(x^{(i)}))) \right] \quad (4)$$

where $\mathbb{E}_{q(\tilde{x}|x)} [\cdot]$ is the expectation over corrupted examples \tilde{x} drawn from the corruption process $q(\tilde{x}|x)$.

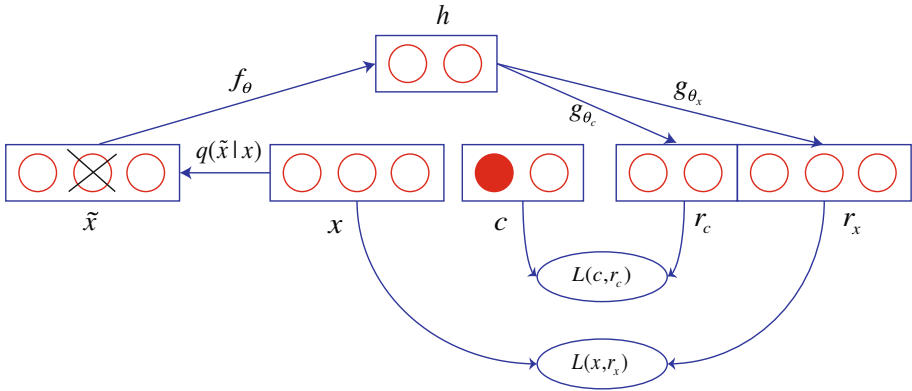


Fig. 2. The architecture of the denoising autoencoder after adding category constraint. c is a standard unit vector, indicating the category of the video where the frame belongs. The hidden layer h attempts to reconstruct x and c together, producing the reconstruction vector r_x and r_c . The objective error function is $L(c, r_c) + \lambda L(x, r_x)$.

The reason why DAE can denoise corrupted data is that the training data usually concentrate near a lower-dimensional manifold, yet most of the time the corruption vector is orthogonal to the manifold. The model learns to project the corrupted data back onto the manifold, thus denoising the input.

3.2 Adding the Category Constraint

Though the features learned by the denoised autoencoders can be highly expressive, as we use the frame-level joint features as the input, all the temporal and category information are discarded. Merely using the model mentioned above, the unsupervised learned features cannot distinguish the significant small differences between similar actions. We modify the denoising autoencoders, adding the category constraint, to make the model capable of emphasizing the imparities in different actions.

Fig. 2 demonstrates our modified autoencoder. Based on the structure of denoising autoencoders, we add an extra target c to the network where c is a vector whose length equals to the action class number d_c . The vector c has only one nonzero element whose index indicates the action type of the video where the example frame belongs. In consequence, a category vector r_c has to be reconstructed by the hidden layer h using a new mapping function g_{θ_c} . Similarly, r_x is the reconstruction vector of x by the mapping function g_{θ_x} . The new training objective of the denoised autoencoder with category constraint (DAE_CC) is:

$$L_{DAE_CC}(\theta) = \sum_i \mathbb{E}_{q(\tilde{x}|x^{(i)})} \left[L(x^{(i)}, g_{\theta_x}(f_{\theta}(x^{(i)}))) + \lambda L(c^{(i)}, g_{\theta_c}(f_{\theta}(x^{(i)}))) \right] \quad (5)$$

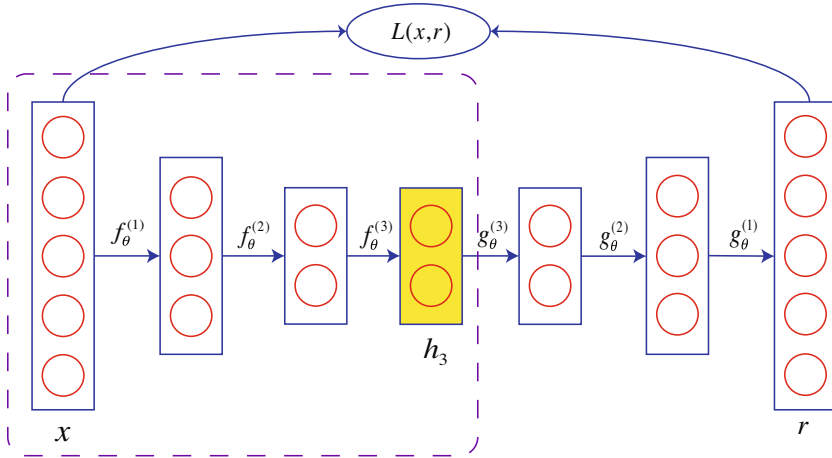


Fig. 3. Fine tuning of the stacking architecture. Each layer autoencoder is trained successively to obtain the encoding and decoding functions, which are used to initialize the parameters of the stacking architecture. All parameters are fine tuning to minimize the reconstruction error $L(x, r)$, by performing gradient descent. The structure inside the dotted box is the model to extract features and the deepest hidden layer h_3 is the final representation we seek.

where λ is a hyper-parameter controlling the strength of the category regularization. It can be optimized by stochastic gradient descent, analogous to the process of optimizing traditional autoencoders.

The reason why we use a regularization term rather than directly learn the class labels as targets is that the input is the joint vector for one frame, yet the class labels are for the whole video. Apparently there are some similar postures among actions. For example, the *stand and put the hands down* posture appears at the beginning of almost all actions. Training the same posture for different labels will lead to trivial results. The regularization term establishes a trade-off between preserving category information and reconstructing the input data.

3.3 Stacked Architecture

By stacking several layers of denoising autoencoders with the category constraint, we build a deep neural network with great expressive power. Greedy layer-wise training is employed: we first train the first layer to get the encoding function f_{θ_1} , then apply it on the clean input to compute the output value, which is used to train the second layer autoencoder to learn f_{θ_2} . The process is repeated from there. At last we fine-tune the deep neural network as in Fig. 3. We use the output of the last autoencoder as the output for the stacked architecture.

3.4 Temporal Representation and Classification

To add temporal information, a temporal pyramid matching (TPM) [6] is used to represent the temporal dynamics of these features. Motivated by Spatial Pyramid Matching (SPM) [21], a max pooling function is used to generate the multi-scale structure. We recursively partition the video sequence into increasingly finer segments along the temporal direction and use max pooling to generate histograms from each sub-region. Typically, 4 levels with each containing 1, 2, 4 and 8 segments are used. The final feature is the concatenation of histograms from all segments.

After the final representation for each video is obtained, a multi-class linear SVM [22] is used to speed up the training and testing, results will be discussed in the next section.

4 Experimental Results

We evaluate our algorithm on a depth-based action recognition dataset, MSR Action3D dataset [12]. We compare our algorithm with several state-of-the-art methods on this dataset, the experimental result shows that our algorithm outperforms these methods. We also reveal the strong denoising capability of our method to reconstruct noisy 3D joint sequences. In all experiments, we train a deep architecture stacking by two autoencoders, where the first one contains 200 nodes in the hidden layer and the second one contains 400 nodes in the hidden layer. We penalize the average output \bar{h}_j of the second autoencoder and pushing it to 0.1, in order to add some sparsity to the model and learn an over-completed representation of joint features. The parameter λ is set to 1.5.

4.1 MSR Action3D Dataset

MSR Action3D dataset [12] is an action dataset of depth sequences captured by a depth camera. The dataset contains 20 actions: *high arm wave, horizontal arm wave, hammer, hand catch, forward punch, high throw, draw x, draw tick, draw circle, hand clap, two hand wave, sideboxing, bend, forward kick, side kick, jogging, tennis swing, tennis serve, golf swing, pick up & throw*. Each action is performed by 10 subjects for three times. There are 567 depth map sequences in total. The provided skeleton data is used to train and test our model. We use the same experimental setting as in [2], half of the subjects are used for training and the rest half for testing. We compare our algorithm with several recent methods and report the results on Table 1. We obtain a recognition accuracy of 87.4%. Fig. 4 shows the confusion matrix of the proposed method. Fig. 5 compares the recognition accuracy for each action of our stacked denoising autoencoders with and without the category constraint. The recognition rate improve from 83.3% to 87.4% after adding the category constraint.

Table 1. Comparison of recognition rate on MSR Action3D Dataset

Method	Accuracy
Recurrent Neural Network [23]	0.425
Dynamic Temporal Warping [24]	0.54
Hidden Markov Model [14]	0.63
STIP [9] + BOW	0.696
Action Graph on Bag of 3D Points [12]	0.747
Eigenjoints [8]	0.823
Random Occupy Pattern [25]	0.865
HON4D [5]	0.859
Proposed Method	0.874

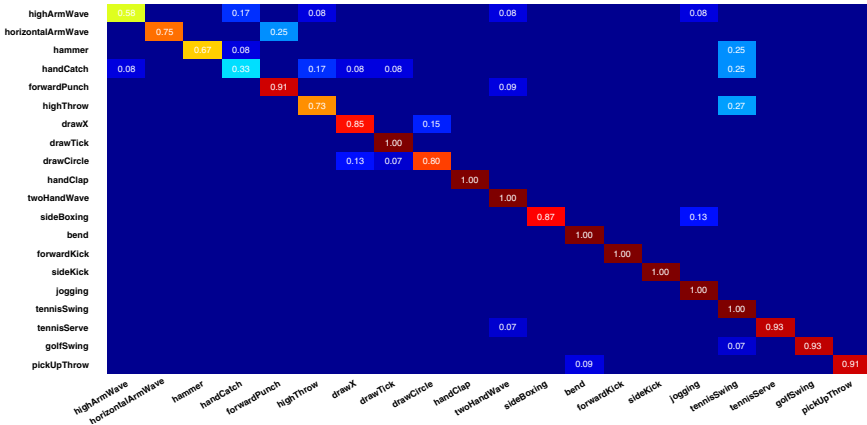


Fig. 4. Confusion matrix of the proposed method on MSR Action3D dataset

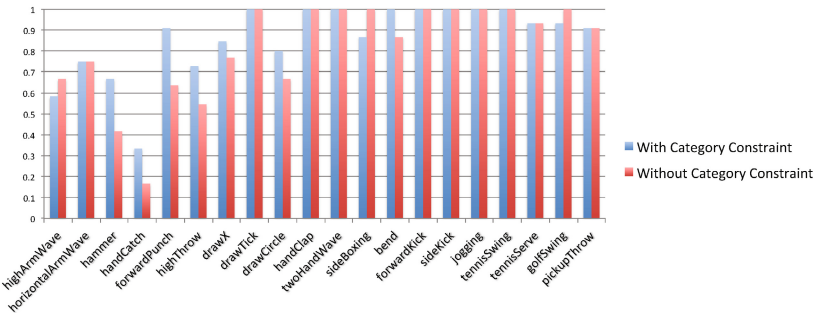


Fig. 5. Comparison of the recognition accuracy for each action before and after adding the category constraint

4.2 Capability to Denoise Corrupted Data

Our model has strong capability to reconstruct realistic data from corrupted input. The top row of Fig. 6 is an action sequence *high arm wave* selected from MSR Action3D dataset. In order to better demonstrate our algorithm efficiency, we add some Gaussian noise to the joint positions and leave out joints stochastically. The bottom row is the reconstruction action sequence, where we can observe that the missing joints are all restored via our model and the motions are more natural and fluent than before.

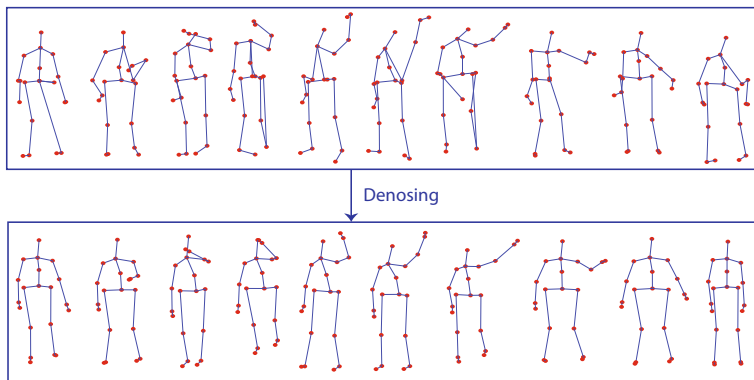


Fig. 6. Examples showing the capability of our model to denoise corrupted data. Top: the corrupted input 3D joint sequence *high arm wave* from MSR Action3D dataset. Bottom: the reconstructed 3D joint sequence.

5 Conclusion

This paper presented a novel feature learning methodology for human action recognition with depth cameras. To better represent the 3D joint features, a deep stacked denoising autoencoder that incorporated with the category constraint was proposed. The proposed model is capable of capturing subtle spatio-temporal details between actions and robust to the noises and errors in the joint positions. The experiments demonstrated the effectiveness and robustness of the proposed approach. In the future, we aim to integrate the temporal information into our feature learning architecture.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (Grant No. 61272317) and the General Program of Natural Science Foundation of Anhui of China (Grant No. 1208085MF90).

References

1. Yang, X., Zhang, C., Tian, Y.: Recognizing actions using depth motion maps-based histograms of oriented gradients. In: Proceedings of the 20th ACM International Conference on Multimedia, pp. 1057–1060. ACM (2012)
2. Wang, J., Liu, Z., Wu, Y., Yuan, J.: Mining actionlet ensemble for action recognition with depth cameras. In: Computer Vision and Pattern Recognition, CVPR (2012)
3. Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M., Moore, R.: Real-time human pose recognition in parts from single depth images. *Communications of the ACM* 56(1), 116–124 (2013)
4. Xia, L., Aggarwal, J.: Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera (2013)
5. Oreifej, O., Liu, Z., Redmond, W.: Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences. In: Computer Vision and Pattern Recognition, CVPR (2013)
6. Luo, J., Wang, W., Qi, H.: Group sparsity and geometry constrained dictionary learning for action recognition from depth maps (2013)
7. Xia, L., Chen, C.C., Aggarwal, J.: View invariant human action recognition using histograms of 3d joints. In: Computer Vision and Pattern Recognition Workshops, CVPRW (2012)
8. Yang, X., Tian, Y.: Eigenjoints-based action recognition using naive-bayes-nearest-neighbor. In: Computer Vision and Pattern Recognition Workshops, CVPRW (2012)
9. Laptev, I.: On space-time interest points. *International Journal of Computer Vision* 64(2-3), 107–123 (2005)
10. Dollár, P., Rabaud, V., Cottrell, G., Belongie, S.: Behavior recognition via sparse spatio-temporal features. In: 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance 2005, pp. 65–72. IEEE (2005)
11. Laptev, I., Marszalek, M., Schmid, C., Rozenfeld, B.: Learning realistic human actions from movies. In: Computer Vision and Pattern Recognition, CVPR (2008)
12. Li, W., Zhang, Z., Liu, Z.: Action recognition based on a bag of 3d points. In: Computer Vision and Pattern Recognition Workshops, CVPRW (2010)
13. Scholkopf, B., Mullert, K.R.: Fisher discriminant analysis with kernels. *Neural Networks for Signal Processing IX*
14. Rabiner, L.: A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77(2), 257–286 (1989)
15. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. *Science* 313(5786), 504–507 (2006)
16. Hinton, G.E., Osindero, S., Teh, Y.W.: A fast learning algorithm for deep belief nets. *Neural Computation* 18(7), 1527–1554 (2006)
17. Bengio, Y.: Learning deep architectures for ai. *Foundations and Trends® in Machine Learning* 2(1), 1–127 (2009)
18. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11), 2278–2324 (1998)
19. Bengio, Y., Courville, A., Vincent, P.: Representation learning: A review and new perspectives (2013)

20. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A.: Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *The Journal of Machine Learning Research* 9999, 3371–3408 (2010)
21. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: *Computer Vision and Pattern Recognition, CVPR* (2006)
22. Chang, C.C., Lin, C.J.: Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2(3), 27 (2011)
23. Martens, J., Sutskever, I.: Learning recurrent neural networks with hessian-free optimization. In: *Proceedings of the 28th International Conference on Machine Learning (ICML 2011)*, pp. 1033–1040 (2011)
24. Müller, M., Röder, T.: Motion templates for automatic classification and retrieval of motion capture data. In: *Proceedings of the 2006 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 137–146. Eurographics Association (2006)
25. Wang, J., Liu, Z., Chorowski, J., Chen, Z., Wu, Y.: Robust 3d action recognition with random occupancy patterns. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) *ECCV 2012, Part II. LNCS*, vol. 7573, pp. 872–885. Springer, Heidelberg (2012)

Author Index

- Basterrech, Sebastián 475
Buriánek, Tomáš 475
- Chang, Chi-Kao 123
Chang, Shih-Hao 287
Chen, Chao-Ho 343
Chen, Chien-Ming 31, 145
Chen, Chih-Min 343
Chen, Chun-Hao 69
Chen, Der-Fa 279
Chen, Jui-Le 97
Chen, S.H. 103
Chen, Shang-Liang 79
Chen, Shi-Huang 413
Chen, Shuo-Tsung 279
Chen, Tsong-Yi 343
Chen, Y.H. 103
Chen, Yi-Ting 451
Chen, Yu-Ming 123
Chen, Yun-Yao 79
Cheng, E-Liang 403
Cheng, Shin-Ming 287
Cheng, Yuh-Ming 403
Cheung, Simon K.S. 11
Chuansheng, Zhou 189
Chung, Chih-Chao 391
Chung, Kai-Lun 465
Chung, Tsui-Ping 145
Cui, Guangzhao 245
- Ding, Hui 57
Ding, Qun 267, 431, 485, 505
Du, Youfu 329
- Fan, Chunlei 431
- Frnda, Jaroslav 165
Fu, Xiaoyan 57
- Gan, Wensheng 135
Guo, Jun-Yi 465
Guo, Xiaotang 299
- He, Bing-Zhe 155
Hong, Tzung-Pei 69, 87, 135
Horng, Mong-Fong 451
Hou, Yi 175, 515
Hsieh, Ching-Hsun 201
Hsu, Chiang 79
Hsu, Hung-Chuan 87
Hsu, Wei-Lin 367
Hu, Jen-Wei 97
Hu, Wu-Chih 343
Huang, Hsiang-Cheh 465
Huang, Hsiu-Chu 391
Huang, Huang-Nan 279
Huang, Yu-Chian 213
- Jhou, Jheng-Syu 413
Jiang, Ji-Han 355
Jin, Peiquan 521
- Kuan, Chie-Yang 155
Kudelka, Milos 47
Kuo, Chang-Ming 123
Kuo, Chung-Ming 123
Kuo, Sheng-Huang 403
Kuo, Su-Hui 441
- Lan, Guo-Cheng 69
Lee, Chung-Hong 213

- Li, Ci-Rong 155
 Li, Fenglian 3
 Li, Jin 181
 Li, Xiaomei 495
 Li, Zhiqiang 431
 Liao, Bin-Yih 451
 Lin, Chiu-Chun 279
 Lin, Chun-Wei 87, 135
 Lin, Yui-Kai 69
 Liu, Guangyu 3
 Liu, Songlin 323
 Liu, Songyan 431
 Liu, Tenghong 323
 Lou, Shi-Jer 391
 Lu, Guangming 299
 Lu, Yuh-Yih 465
- Ma, Jialin 175, 515
 Ma, Zhaozhe 495
 Meng, Lei 317
 Miaoyan, Li 189
 Mpountouropoulos, Nikolaos 309
 Muchao, Lu 21
- Nikolaidis, Nikos 309
- Ouyang, Zhengzheng 223
- Pan, Jeng-Shyang 31, 451
 Papachristou, Konstantinos 113
 Pitas, Ioannis 113, 309
 Platos, Jan 47
- Sang, Ruoxin 521
 Sevcik, Lukas 165
 Shang, Yuanyuan 57
 Shie, Shih-Chieh 355
 Shyu, Fong-Ming 235
 Snášel, Václav 245
 Song, Bingbing 267
 Song, Bo 181, 495
 Su, Jingyong 39
 Su, Yi-Jen 367
 Sun, Hung-Min 155
- Tang, Lin-Lin 39
 Tang, Xiaoyue 223
 Tefas, Anastasios 113, 309
- Tsai, Chang-Ling 377
 Tsai, Huei-Yin 391
 Tsai, Jr-Yung 355
 Tsai, Ming-Te 451
 Tseng, Kuo-Kun 287
 Tso, Raylin 135
- Voznak, Miroslav 165
- Wan, Shouhong 521
 Wang, Chia-Hui 201
 Wang, Chuen-Ching 383
 Wang, Erfu 505
 Wang, Eric Ke 31, 145
 Wang, Hsuan-Pei 79
 Wang, Hui 223
 Wang, Shyue-Liang 69
 Wang, Song-Nian 235
 Wang, Tien-Chin 441
 Wang, Yazhuo 299
 Wang, Yih-Fuh 377
 Wei, Chi-Hung 383
 Weng, Chi-Yao 155
 Wu, Chun-Yen 213
 Wu, Jie 245
 Wu, Tsu-Yang 31, 145
 Wu, Wu-En 155
 Wun, Jian-Cheng 367
- Yang, Chu-Sing 97
 Yang, Hsin-Chang 213
 Yang, Nai-Chung 123
 Yu, Wei 223
- Zehnalova, Sarka 47
 Zhang, Wei 3
 Zhang, Xueying 3
 Zhao, Bing 485, 505
 Zhao, Dingtao 441
 Zhao, Yingce 299
 Zhao, Yongyi 181
 Zhen, Jiaqi 485, 505
 Zheng, Xinying 145
 Zhou, Chengxiang 31
 Zhou, Chuansheng 317
 Zhou, Rurui 257
 Zhou, Xiuzhuang 57
 Zjavka, Ladislav 421