

Ryszard S. Choraś (Ed.)

Image Processing and Communications Challenges 2

Advances in Intelligent and Soft Computing

Editor-in-Chief

Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Further volumes of this series can be found on our homepage: springer.com

Vol. 70. Y. Demazeau, F. Dignum, J.M. Corchado, J. Bajo Pérez (Eds.)
Advances in Practical Applications of Agents and Multiagent Systems, 2010
ISBN 978-3-642-12383-2

Vol. 71. Y. Demazeau, F. Dignum, J.M. Corchado, J. Bajo, R. Corchuelo, E. Corchado, F. Fernández-Riverola, V.J. Julián, P. Pawlewski, A. Campbell (Eds.)
Trends in Practical Applications of Agents and Multiagent Systems, 2010
ISBN 978-3-642-12432-7

Vol. 72. J.C. Augusto, J.M. Corchado, P. Novais, C. Analide (Eds.)
Ambient Intelligence and Future Trends, 2010
ISBN 978-3-642-13267-4

Vol. 73. J.M. Corchado, P. Novais, C. Analide, J. Sedano (Eds.)
Soft Computing Models in Industrial and Environmental Applications, 5th International Workshop (SOCO 2010), 2010
ISBN 978-3-642-13160-8

Vol. 74. M.P. Rocha, F.F. Riverola, H. Shatkay, J.M. Corchado (Eds.)
Advances in Bioinformatics, 2010
ISBN 978-3-642-13213-1

Vol. 75. X.Z. Gao, A. Gaspar-Cunha, M. Köppen, G. Schaefer, and J. Wang (Eds.)
Soft Computing in Industrial Applications, 2010
ISBN 978-3-642-11281-2

Vol. 76. T. Bastiaens, U. Baumöl, and B.J. Krämer (Eds.)
On Collective Intelligence, 2010
ISBN 978-3-642-14480-6

Vol. 77. C. Borgelt, G. González-Rodríguez, W. Trutschnig, M.A. Lubiano, M.Á. Gil, P. Grzegorzewski, and O. Hryniewicz (Eds.)
Combining Soft Computing and Statistical Methods in Data Analysis, 2010
ISBN 978-3-642-14745-6

Vol. 78. B.-Y. Cao, G.-J. Wang, S.-Z. Guo, and S.-L. Chen (Eds.)
Fuzzy Information and Engineering 2010
ISBN 978-3-642-14879-8

Vol. 79. A.P. de Leon F. de Carvalho, S. Rodríguez-González, J.F. De Paz Santana, and J.M. Corchado Rodríguez (Eds.)
Distributed Computing and Artificial Intelligence, 2010
ISBN 978-3-642-14882-8

Vol. 80. N.T. Nguyen, A. Zgrzywa, and A. Czyzewski (Eds.)
Advances in Multimedia and Network Information System Technologies, 2010
ISBN 978-3-642-14988-7

Vol. 81. J. Düh, H. Hufnagl, E. Juritsch, R. Pfliegl, H.-K. Schimany, and Hans Schönegger (Eds.)
Data and Mobility, 2010
ISBN 978-3-642-15502-4

Vol. 82. B.-Y. Cao, G.-J. Wang, S.-L. Chen, and S.-Z. Guo (Eds.)
Quantitative Logic and Soft Computing 2010
ISBN 978-3-642-15659-5

Vol. 83.xxx

Vol. 84. Ryszard S. Choraś (Ed.)
Image Processing and Communications Challenges 2
ISBN 978-3-642-16294-7

Ryszard S. Choraś (Ed.)

Image Processing and Communications Challenges 2

Editors

Prof. Ryszard S. Choraś
University of Technology and Life Sciences
Kordeckiego 20
85-225 Bydgoszcz
Poland
E-mail: choras@utp.edu.pl

ISBN 978-3-642-16294-7

e-ISBN 978-3-642-16295-4

DOI 10.1007/978-3-642-16295-4

Advances in Intelligent and Soft Computing

ISSN 1867-5662

Library of Congress Control Number: 2010936133

© 2010 Springer-Verlag Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable for prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, 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.

Typeset & Cover Design: Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed on acid-free paper

5 4 3 2 1 0

springer.com

Foreword

Image Processing and Communications represents an exciting and dynamic part of the information area.

This book consists of 52 scientific and technical papers from 14 Nations, after a careful selection performed by many international reviewers. The papers are conveniently grouped into 6 chapters:

- Computer Vision and Image Processing
- Biometric
- Recognition and Classification
- Biomedical Image Processing
- Applications
- Communications.

Each chapter focuses on a specific topic, presents results, and points out challenges and future directions.

A number of people helped in making this book a reality. Adam Marchewka, PhD. demonstrated tremendous commitment in typesetting the book. I thank Karolina Skowron, who typed and corrected some portion of the text. I express my gratitude to all reviewers. I am also grateful to Thomas Ditzinger, editor at Springer for keeping me on schedule for the production of the book.

Bydgoszcz,
October 2010

Ryszard S. Choraś

Contents

Part I: Computer Vision and Image Processing

Earliest Computer Vision Systems in Poland	3
<i>Ryszard Tadeusiewicz</i>	
VICAL: Visual Cognitive Architecture for Concepts Learning to Understanding Semantic Image Content	15
<i>Yamina Mohamed Ben Ali</i>	
Implementation of Computer Vision Algorithms in DirectShow Technology	31
<i>Piotr Szczypiński, Paweł Pełczyński, Dominik Szajerman, and Paweł Strumiłło</i>	
Implementation of Hurwitz-Radon Matrices in Shape Representation	39
<i>Dariusz Jakóbczak</i>	
Video Quality Assessment Using the Combined Full-Reference Approach	51
<i>Krzysztof Okarma</i>	
An Improved Self-embedding Algorithm: Digital Content Protection against Compression Attacks in Digital Watermarking	59
<i>Pratheepan Yogarajah, Joan Condell, Kevin Curran, Abbas Cheddad, and Paul McKeiv</i>	
Generation of View Representation from View Points on Spiral Trajectory	67
<i>Andrzej Salamończyk and Wojciech Mokrzycki</i>	

Gradient Based Edge Detection in Various Color Spaces	75
<i>Marek A. Samko</i>	

Improve Vector Quantization Strategy	81
<i>Zahraa F. Muhsen, Loay A. Jorj, and Imad H. Alhussaini</i>	

Part II: Biometric

Knuckle Biometrics for Human Identification	91
<i>Michał Choraś and Rafał Kozik</i>	

A New Method of Fingerprint Key Protection of Grid Credential	99
<i>Yarema Varetsky, Bogdan Rusyn, Agnieszka Molga, and Anatolii Ignatovych</i>	

Human Vein Pattern Segmentation from Low Quality Images – A Comparison of Methods	105
<i>Rafał Kabaciński and Mateusz Kowalski</i>	

A Modified Algorithm for User Identification by His Typing on the Keyboard	113
<i>Piotr Panasiuk and Khalid Saeed</i>	

Multimodal Biometric Personal Authentication Integrating Iris and Retina Images	121
<i>Ryszard S. Choraś</i>	

Part III: Recognition and Classification

Fusion Methods for the Two Class Recognition Problem – Analytical and Experimental Results	135
<i>Michał Woźniak and Marcin Zmyślony</i>	

Feature Type and Size Selection for AdaBoost Face Detection Algorithm	143
<i>Jerzy Dembski</i>	

3D Morphable Models Application for Expanding Face Database Limited to Single Frontal Face Image Per Person	151
<i>Lukasz Kulasek and Andrzej Czyżewski</i>	

A Partition of Feature Space Based on Information Energy in Classification with Fuzzy Observations	159
<i>Robert Burduk</i>	
Recognition of Signed Expressions Using Cluster-Based Segmentation of Time Series	167
<i>Mariusz Oszust and Marian Wysocki</i>	
Extending 3D Shape Measurement with Reflectance Estimation	175
<i>Robert Sitnik, Jakub Krzesłowski, and Grzegorz Mączkowski</i>	
Software Framework for Efficient Tensor Representation and Decompositions for Pattern Recognition in Computer Vision	185
<i>Bogusław Cyganek</i>	
Hand Shape Recognition in Real Images Using Hierarchical Temporal Memory Trained on Synthetic Data	193
<i>Tomasz Kapuściński</i>	
Performance Comparison among Complex Wavelet Transforms Based Face Recognition Systems	201
<i>Alaa Eleyan and Hasan Demirel</i>	

Part IV: Biomedical Image Processing

The Method of Immunohistochemical Images Standardization	213
<i>Anna Korzyńska, Urszula Neuman, Carlos Lopez, Marylene Lejeun, and Ramon Bosch</i>	
The Usefulness of Textural Features in Prostate Cancer Diagnosis	223
<i>Jacek Śmietański</i>	
Noise Influence Reduction in Estimation of CBF, CBV and MTT, MRI Perfusion Parameters	231
<i>Rafał Henryk Kartaszyński and Paweł Mikołajczak</i>	
Interpretation of the Sequences of Magnetocardiographical Images Based on Flow of Electrical Impulses through Human Heart	239
<i>Kamila Baron-Pałucka</i>	

Automatic Left Ventricle Segmentation in T2 Weighted CMR Images	247
<i>Kurhairy Abdul Kadir, A. Payne, John J. Soraghan, and C. Berry</i>	
Research of Muscular Activity during Gait of Persons with Cerebral Palsy	255
<i>Kristina Daunoravičienė, Jolanta Pauk, Jim Raso, and Julius Griškevičius</i>	
Automated Recognition of Abnormal Structures in WCE Images Based on Texture Most Discriminative Descriptors	263
<i>Piotr Szczypiński and Artur Klepaczko</i>	
Augmented Reality Interface for Visualization of Volumetric Medical Data	271
<i>Tomasz Hachaj and Marek R. Ogiela</i>	
Biomedical Computer Vision Using Computer Algebra: Analysis of a Case of Rhinocerebral Mucormycosis in a Diabetic Boy	279
<i>Mario Vélez and Juan Ospina</i>	

Part V: Applications

Adaptive B-Spline Model Based Probabilistic Active Contour for Weld Defect Detection in Radiographic Imaging	289
<i>Nafaa Nacereddine, Latifa Hamami, Djemel Ziou, and Aicha Baya Goumeidane</i>	
FONN-Based Affine-Invariant Image Recognition	299
<i>Bartłomiej Stasiak</i>	
Coarse-Grained Loop Parallelization for Image Processing and Communication Applications	307
<i>Włodzimierz Bielecki and Marek Palkowski</i>	
SMAS - Stereovision Mobility Aid System for People with a Vision Impairment	315
<i>Rafał Kozik</i>	
Extracting Symbolic Function Expressions by Means of Neural Networks	323
<i>Jarostaw Majewski and Ryszard Wojtyna</i>	

Mathematical Morphology in the Process of Musical Notation Recognition	331
<i>Arkadiusz Rajs</i>	
GPU-Accelerated Object Tracking Using Particle Filtering and Appearance-Adaptive Models	337
<i>Bogusław Rymut and Bogdan Kwolek</i>	
Application of Epipolar Rectification Algorithm in 3D Television	345
<i>Jakub Stankowski and Krzysztof Klimaszewski</i>	
Crack Detection on Asphalt Surface Image Using Local Minimum Analysis	353
<i>Adam Marchewka</i>	
Eye Tracking System for Human Computer Interaction	361
<i>Mscisław Śrutek and Łukasz Matuszak</i>	

Part VI: Communications

Errors Nature in Indoors Low Power 433 MHz Wireless Network	373
<i>Bartosz Boryna, Bożydar Dubalski, Piotr Kiedrowski, and Antoni Zabłudowski</i>	
Using Google Earth for Visualization in FTTH Network Planning	379
<i>Jens Myrup Pedersen, Gustav Helgi Haraldsson, and M. Tahir Riaz</i>	
The Development of a Platform Based on Wireless Sensors Network and ZigBee Protocol for the Easy Detection of the Forest Fire. A Case Study	391
<i>Andreas Vlissidis, Stavros Charakopoulos, and Emmanouil Makrygiannakis</i>	
Mazovia Broadband Network (MBN Network). Case Study	401
<i>Antoni Zabłudowski, Bożydar Dubalski, and Łukasz Zabłudowski</i>	
The Method of GMPLS Network Reliability Evaluation	417
<i>Janusz Korniak and Paweł Różycki</i>	
The Improved Least Interference Routing Algorithm	425
<i>Ireneusz Olszewski</i>	

Comparison of Modified Degree 6 Chordal Rings	435
<i>Sławomir Bujnowski, Bożydar Dubalski, Antoni Zabłudowski, Damian Ledźński, Tomasz Marciniak, and Jens. M. Pedersen</i>	
Evaluation of Measurement Based Admission Control Algorithms for IEEE 802.16 Networks in Simulations with L2S Physical Layer Abstraction and nbLDPC Codes	447
<i>Adam Flizikowski, Marcin Przybyszewski, and Witold Hohubowicz</i>	
The Gap between Packet Level QoS and Objective QoE Assessment of WWW on Mobile Devices	461
<i>Adam Flizikowski, Damian Puchalski, Kacper Sachajdak, and Witold Hohubowicz</i>	
Evaluation of Smoothing Algorithms for a RSSI-Based Device-Free Passive Localisation	469
<i>Gabriel Deak, Kevin Curran, and Joan Condell</i>	
Performance Evaluation of ADS System Based on Redundant Dictionary	477
<i>Rafał Renk, Lukasz Saganowski, Michał Choraś, and Witold Hohubowicz</i>	
Author Index	485

Computer Vision and Image Processing

Earliest Computer Vision Systems in Poland

Ryszard Tadeusiewicz

AGH University of Science and Technology,
Faculty of Electrical Engineering, Automatic Control,
Computer Science and Electronics, Department of Automatic Control
e-mail: rtad@agh.edu.pl

Summary. In the paper is presented the brief history of research conducted in Poland (at AGH University) in the area of image processing, analysis and recognition. The history is connected with changes of the technology used for problem solving. First systems are build from the separate integrated circuits of low and medium scale of integration, after this we develop some systems build on the base of VLSI elements, and now the new systems are constructed on the base of FPGA technology. Nevertheless all the time scientific group from AGH was conducted research in technological applications of vision systems. At 70th and 80th years of XX century it was pioneer works devoted to building of second in Poland and one of the first in Europe systems for computer processing of the images, now research is dedicated mainly toward real time image processing systems.

1 Introduction

The computer vision technique is currently very popular and commonly used. It is enough to mention omnipresent digital cameras and DVCs which were so technically improved and are produced so economically that they are universally installed in mobile phones as a bonus. Digital images also accompany us in various forms of *multimedia communication* (vide [\[1\]](#)) and are the main content of data resources - e.g. the Internet. Therefore, for present-day computer user image manipulating is something very natural and obvious.

However, we should realize that it was not always like that! Digital technique, as the name indicates, was originally created to operate on numbers not images, and the computer - also as the name suggests - is a machine used for calculations and not determining liver deformations on sonogram. The fact that we can so freely use the image processing and recognition techniques is a result of long-term process of techniques development and improvement and AGH University staff also participated in this historical process. In this article I would like to recall some of our achievements in this field and show the long and crooked path that we had to walk to reach our current level of competence, related to methods defined often as *computer vision*.

Usually, when we publish articles concerning computer vision, we refer to recent achievements in this field (vide [2, 3, 4]) but in this case, as exception to the rule, I will intentionally go back to the past. I will try to present the beginning of computer vision development at AGH University in the "pioneer" stage of this technique development, when very few people knew what was this computer vision about and many authorities claimed that computers should not be used for such activities because that is not their purpose. AGH is a university of science and technology therefore industry needs, especially fields related to mining and metallurgy, were always the priority in the computer vision techniques development. It is worth mentioning that the first *vision system* constructed at AGH University was used to control the process of mineral resources enrichment, more precisely the technology of copper, zinc and lead ore enrichment on the basis of froth image observed by CCTV camera and processed by a computer modified to working in industrial conditions. Further research was performed in order to provide industrial robots carrying out simple inspection works, assembling (including welding) and sorting out objects with vision feedback.

2 The Problem of Computer Image Input

It is very difficult for a present-day user of omnipresent and highly efficient devices, that allow quick and convenient ways of obtaining digital images and inputting them into computer, to imagine how big problem that was in the past. When we started the research into automatic image processing and analysis at AGH University, the main method of computer data input were punched tapes and cards (Fig. 1).

First, we had to design, work out and then personally build and test was electronic devices that allowed image processing and its direct input to the computer. It seems that first publications about issues that are considered here were published in mid 70s and many following publications in 1978 were the result of them (vide [7, 8, 9]). As I mentioned, the main problem at that time was the converter that allowed an input of images supplied at high speed by typical (analogue) TV cameras into slow, at that time, computers. Constructing such converter using electronic systems available in 70s was very difficult and moreover the *stream of digital information* produced by such converter was too big and fast for capabilities of computers then. Therefore, it was necessary to use buffer memory that intercepted digital image - of course this memory was also constructed personally from electronic systems of very small (for present requirements) integration scale.

However, having specific objective, which was the automation of process of mineral resources enrichment in mining industry (vide [8]) and practically unlimited (thanks to cooperation with KGHM company) financial resources, because financing research concerning mining was the priority in the mid 70s, we built the computer image processing system, called CESARO (Cyfrowy



Fig. 1. Old methods of data input were not favourable to image processing techniques

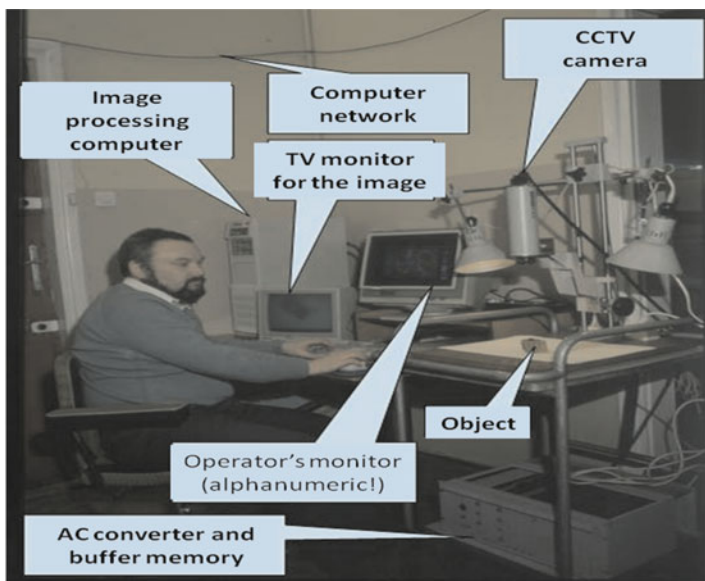


Fig. 2. The first CESARO vision system constructed at AGH University

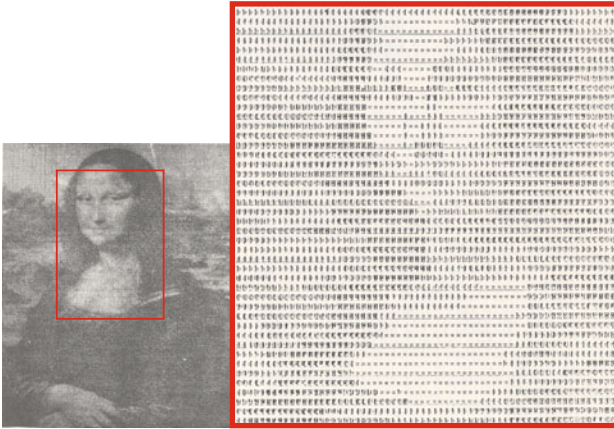


Fig. 3. The method of presenting a digital image using alphanumeric printer

Eksperymentalny System Analizy i Rozpoznawania Obrazów [Experimental Digital System of Analysis and Image Recognition]- see Fig. 2).

It was the second in Poland, and one of the first in Europe system of that type. CESARO was frequently described [9, 10] and became the basis of various research in the field of computer image analysis [11, 12]. It is worth emphasising, that during work on construction and software of the first Polish vision systems we encountered many difficulties and problems, that modern computer vision researchers do not even realise that they existed. Not only the image input was very difficult but also it was very hard to see the image once it was in a machine memory - for example, after performing an operation that was part of its processing and analysis. Nowadays it is hard to imagine, but monitors and printers used at that time did not have the capability of displaying or printing images because they were purely alphanumeric. We could manage with displaying an image by connecting (through digital-analogue converter) typical studio monitor (used in broadcasting studio to watch broadcasted program) to computer. Appropriate software was scanning its memory that included image that we were interested in (for example, the result of processing) and emitted signals that controlled work of converter displaying image on screen. Of course, we had many problems concerning details (eg. synchronization computer with monitor) but we managed to solve them.

The real difficulty occurred when it was necessary to get *lasting copy* of processed or analysed digital image. As opposed to monitor, which electronic construction was reasonably *modification-friendly*, printers contained many indispensable mechanical elements and their production or modification at AGH University was quite difficult. Precision engineering (or rather lack of it...) is until present the limiting factor of our University capabilities in many interesting research fields - e.g. microrobotics.

Back to the image printing problem. It turned out that the only chance to get "image-like" recording on paper is the technique of typing *printing characters* on each other in such a way that the summary degree of blackening reached desirable value. Fig. 3 depicts an exemplary printing obtained by this method.

For people alarmed by the context of this exemplary image, I would like to make clear that AGH staff WAS NOT engaged in processing such images but there is no alphanumerically generated scientific image sample left. Fortunately, I managed to find the example presented above of printing of another (famous) image.

3 Previous Research Related to Computer Vision

CESARO was unique with regard to its range of application (e.g. it was used to optimise copper ore flotation) and was the base of technological studies, which were widely discussed among specialists related to mineral resources processing [10, 13, 14]. After years of using (constantly developed) CESARO system and designing new algorithms of computer vision and programming them in a way that allowed high processing speed despite huge data set which represents every image in *information system*, in 1982 we managed to present the first extensive methodological study describing capabilities and limitations of image techniques, which were fairly unknown novelty at that time [15]. Well-known version of this study, published in scientific journal *Informatyka* [16] was for many following years the inspiration for numerous attempts to construct vision systems in various research centres in Poland (it is confirmed by citations of mentioned article in other researchers' studies). The biggest advantage of CESARO system was the fact that despite highly advanced technological level (of course in relation to science and technology condition when it was created and developed) it was built entirely from **domestic** components that were easy to obtain and relatively cheap. Modern researchers would probably find this criterion odd and absurd, because nowadays we are trying to select components to build given research station according to its quality ('only the best are good enough') and not place where they were made. Market became global. 80s were different, though. It was very hard (and sometimes impossible) to get imported elements and components and rates of exchange were so high that even the cheapest elements from Western Countries were much more expensive than their equivalents produced in Eastern Bloc. Therefore, during presentations of CESARO systems at various conferences we emphasised its construction based on domestic components, because it was significant at that time [17]. Especially the comparison of functionality and efficiency between our system and other researchers' systems gave rather encouraging results [18].

After constructing the CESARO system we did not confine ourselves to its activation and using it for practical purposes. We also searched for method that would allow higher image processing speed on the basis of fact that a

packed image was located in computer memory in such a way that one cell contained a few pixels (vide [9]). Again a short comment: at that time byte structure was not the only one in use, as it is nowadays, but also memory cell was used as a unit processed in one processor working cycle. This technique of speeding up calculations may seem a bit exotic from present point of view but it turned out to be surprisingly effective which allowed to increase efficiency of processing without using multiprocessor architectures, which were very expensive and hard to get. We also gained another advantages concerning software by implementing new and original software architecture using hierarchical relations between operations performed on image at various precision levels of description [20]. Nowadays, these ideas may become interesting only as a curiosity but once it was very practicable and important way to gain higher efficiency of processing, even when using very slow computers, that were the only available ones at that time.

Vision system that we constructed encouraged us to research various fields of its usage. Because works were performed mainly at AGH *Department of Automatic Control* - the first field of applications was automatics [21] (e.g. window glass drawing at glassworks [22]), and then attempts of applications in robotics [23, 24, 25]. Later we studied applications in telecommunications [26] and medicine [27] which became the main subject of studies at AGH for a long time.



Fig. 4. Multiprocessor vision system CESARO2

4 Multiprocessor Systems. CESARO2

Tasks entrusted to image processing systems, were becoming more and more serious and thus 10 year old CESARO system, in spite of constant modifications and development, did not meet the requirements anymore - especially *temporal* ones, related to using image technique in control systems and other real time applications. Therefore, at the end of 80s, at AGH *Department of Automatic Control*, we started constructing system CESARO2, whose distinctive feature was *multiprocessing* [28, 29]. New system was based mainly on biocybernetic grounds and used theory and practice of *neural-like* systems that had been developed at AGH for many years [30]. Thanks to popularization of the idea of building multiprocessor vision system for robotics [31] we managed to get necessary resources and built multiprocessor vision system CESARO2 (Fig. 4 depicts a version at AGH Museum of History of Science). This system was used experimentally as a visual feedback for industrial robot IRb6 (ASEA-PIAP), which became a base for a book "Vision systems in industrial robots" [32] which was the first easily accessible monograph that described principles of computer image analysis, processing and recognition along with their practical realization in a form of specialized electronic systems. A book "Image recognition" [33], published a year before, focused on theoretical and algorithmical approach.

Wide possibilities of research on image processing and recognition, using CESARO2 system, at the beginning of 90s called AGH researchers' attention to possibilities that were allowed by neural networks technique, which were becoming more and more popular at that time. A study [34] which was the first (as it seems) Polish review on possibilities at this field was published then, and another study [35] proposed specific methodology of combining computer image processing and analysis with method and techniques of neural networks. Developing this idea and procedure lead to at least dozen of very interesting Ph. D theses across Poland because combination of image processing technique and neural networks turned out to be highly useful instrument in a surprisingly wide range of issues and problems caused by practical needs [36, 37]. An example of fairly unexpected usage of combined image and neural techniques may be a work [38] presenting the possibilities of those techniques in food technology. Because of extensive confirmation of practical usefulness, instrument, which was a combination of image and neural techniques, was a subject of extensive basic research conducted at AGH from 1992 and crowned among other things with a publication [39].

Along with the application of CESARO2 system other works were also performed, e.g. one of the first in Poland research on capabilities of using computer image recognition to read printed texts and handwriting. Results of these studies were the basis of scientific papers as well as popular studies [40, 41], which presented the capabilities of this (fairly unknown at that time) technique to Polish IT specialists. The result of these studies was e.g. practical solution that used optical reading technique for a fast input of handwriting

student application forms into enrolment system at AGH, and also interesting scientific results [42].

5 And Then There Was an Avalanche ...

The review of pioneer research and constructional studies concerning computer image processing, carried out at AGH from 70s should be finished at mid 90s. Not because there was not any further studies on this subject. On the contrary, in 90s and at the beginning of 21st century many interesting works concerning image processing and recognition were (and still are) published at AGH. Highly advanced research on recognition and understanding of medical images was started (professor Marek Ogiela). Interesting subject of traffic video detectors used in automatic monitoring and traffic control was taken up (professor Andrzej Adamski, Zbigniew Mikrut, Ph.D). There are also studies that aim for realization of increasing amount of tasks assigned for computer vision using specialised processors produced with FPGA technology (Marek Gorgoń, Ph.D). Many interesting problems related to using image processing technique in industrial automatics were solved (Zbigniew Bublński, Ph.D and Piotr Pawlik, Ph.D). A reviews of those further studies can be found in publications [43, 44, 45, 46] and in textbook that was created on their basis [47].

However, despite intensive development of computer image analysis, recognition and understanding at AGH, this field stopped to be distinguishing mark for our university, because now almost everybody are engaged in such studies. At AGH very interesting studies related to image processing are carried out by professor Kazimierz Wiatr, professor Mariusz Ziółko, Przemysław Korohoda, Ph.D, Bogusław Cyganek Ph.D (Electronics Department) and professor Tomasz Zieliński (Department of Measurement and Instrumentation).

Researchers from other centres also work very hard and achieve interesting results (e.g. professor Leszek Wojnar, Cracow University of Technology; professor Konrad Wojciechowski, Silesian University of Technology; professor Marek Kurzyński, Wrocław University of Technology; professor Andrzej Materka, Technical University of Łódź; professor Witold Malina, Gdańsk University of Technology; professor Adam Dąbrowski, Poznań University of Technology - and of course Warsaw research centre lead by professor Juliusz Lech Kulikowski, senior in this field, with many young researchers, who bring knowledge, talent and enthusiasm to these studies (e.g. professor Artur Przelaskowski, Warsaw University of Technology). Therefore, we do not have to worry about the future of computer image analysis, processing and recognition, because the most brilliant minds and the most powerful centres are researching into it. Then from time to time, not too often, but for example on the occasion of anniversary, it is worth reminding the past of this field - which was the purpose of this article.

References

1. Tadeusiewicz, R.: Rola technik cyfrowych w komunikacji społecznej oraz w kulturze i edukacji. In: Handzelewicz, M. (ed.) *Cyfrowy świat bibliotek - problemy techniczne, prawne, wdrożeniowe*, CPI Warszawa, pp. 7–78 (2006)
2. Tadeusiewicz, R., Ogiela, M.R.: Intelligent Recognition in Medical Pattern Understanding and Cognitive Analysis. In: Sarfraz, M. (ed.) *Computer-Aided Intelligent Recognition Techniques and Applications*, pp. 257–274. John Wiley & Sons, Ltd., Hoboken (2005)
3. Tadeusiewicz, R., Ogiela, M.R.: Picture Languages in Automatic Radiological Palm Interpretation. *International Journal of Applied Mathematics and Computer Science* 15(2), 305–312 (2005)
4. Tadeusiewicz, R., Ogiela, M.R.: New Proposition for Intelligent Systems Design, Artificial Understanding of the Images as the Next Step of Advanced Data Analysis After Automatic Classification and Pattern Recognition. In: Kwasińska, H., Paprzycki, M. (eds.) *Intelligent Systems Design and Applications*, pp. 297–300. IEEE Computer Society Press, Los Alamitos (2005)
5. Tadeusiewicz, R.: Rozpoznawanie obrazów. *Informatyka* 12, 15–18 (1978)
6. Tadeusiewicz, R.: Komputerowa analiza przydatności wybranych metod rozpoznawania obrazów w diagnostyce neuroinfekcji. W mat. II Krajowej Konferencji Biocybernetyka i Inżynieria Biomedyczna, Gliwice, pp. 63–64 (1978)
7. Tadeusiewicz, R.: Cyfrowe przetwarzanie obrazów telewizyjnych w celu ich analizy i rozpoznawania. In: W mat. IV Symp. Mpn. Weae. Agh., pp. 5–8 (1978)
8. Kordek, J., Nipl, R., Nowikow, P., Romanowski, J., Szuba, T., Tadeusiewicz, R.: Rozpoznawanie obrazów optycznych w przeróbce surowców mineralnych. In: W mat. Symp, *Automatyzacja Procesów Przeróbki Węgla*, Katowice, pp. 148–154 (1978)
9. Tadeusiewicz, R., Kordek, J., Nipl, R.: Cyfrowy eksperymentalny system analizy i rozpoznawania obrazów CESARO. *Zeszyty Naukowe AGH*, 347–355 (1979)
10. Kordek, J., Nipl, R., Sztaba, K., Tadeusiewicz, R.: CESARO - the digital experimental system of analysis and recognition of images. In: 17th International Symposium on the Application of Computer and Mathematics in the Mineral Industries, Moskwa, pp. 393–398 (1980)
11. Tadeusiewicz, R.: Programy analizy składu ziarnowego materiałów sypkich w systemie CESARO. *Zeszyty Naukowe AGH*, 367–375 (1979)
12. Tadeusiewicz, R., Mikrut, Z.: Komputerowa analiza współczynników kształtu ziaren z wykorzystaniem systemu CESARO. *Zeszyty Naukowe AGH*, 356–365 (1979)
13. Sztaba, K., Nipl, R., Kordek, J., Tadeusiewicz, R., Romanowski, J., Nowikow, P.: Auswertung optischer Aufnahmen bei der Aufbereitung mineralischer Rohstoffe. In: *Physikalische Eigenschaften von Kornmengen und Kornschutungen*, Leipzig, pp. 227–235 (1980)
14. Kordek, J., Nipl, R., Sztaba, K., Tadeusiewicz, R.: Eksperymentalna cyfrowa systema CESARO analiza i rozpoznawania obrazow. In: *Primienienie EWM i matematycznych metod w gornom dielie - Trudy 17-go miedzynarodnego simpozjuma*, Moskwa, pp. 452–456 (1982)
15. Tadeusiewicz, R.: Komputerowa analiza obrazów i jej zastosowania. *Elektrotechnika* 1(2), 85–94 (1982)
16. Tadeusiewicz, R., Pachowicz, P.: CESARO - system analizy i rozpoznawania obrazów wizualnych. *Informatyka* 7/8, 27–29 (1983)

17. Pachowicz, P., Tadeusiewicz, R.: System CESARO jako propozycja wykorzystania krajowego sprzętu informatycznego do analizy i przetwarzania obrazów. In: VI Ogólnopolska Konferencja Biocybernetyki i Inżynierii Biomedycznej, Warszawa, pp. 415–417 (1983)
18. Pachowicz, P., Tadeusiewicz, R.: Porównanie systemu CESARO z wybranymi systemami analizy i rozpoznawania obrazów. *Elektrotechnika* 3(2), 173–182 (1984)
19. Pachowicz, P., Tadeusiewicz, R.: Efektywność zastosowania przetwarzania półrównoległego w sekwencyjnych systemach obrazowych. In: I Krajowa Konferencja Przetwarzanie sygnałów w telekomunikacji, sterowaniu i kontroli, pp. 216–219 (1984)
20. Tadeusiewicz, R.: Hierarchiczny system oprogramowania systemu analizy i rozpoznawania obrazów. In: I Krajowa Konferencja Przetwarzanie sygnałów w telekomunikacji, sterowaniu i kontroli, pp. 220–224 (1984)
21. Pachowicz, P., Tadeusiewicz, R.: Zadania, metody i zastosowania w automatyzacji procesów produkcyjnych komputerowej analizy obrazów. *Zastosowania Komputerów w Przemśle, Szczecin 1983*, 103–114 (1983)
22. Skowiniak, A., Tadeusiewicz, R.: Eine Konzeption des Visionssystems für die Automatisierung der Ziehmaschinen. In: 14 Glastechnikertagung der DDR, Berlin, pp. 42–44, pp. 198–203 (1987)
23. Tadeusiewicz, R.: Rola podsystemu wizyjnego w systemie sterowania robota. *Prace Naukowe Instytutu Cybernetyki Technicznej Wrocław PW 75*, 249–256 (1988)
24. Tadeusiewicz, R.: Systemy wizyjne robotów przemysłowych. *Materiały III Krajowej Konferencji Naukowo - Technicznej Przetwarzanie sygnałów w telekomunikacji, sterowaniu i kontroli, Bydgoszcz 1988*, 1–25 (1988)
25. Tadeusiewicz, R.: Systemy wizyjne dla robotów przemysłowych, rola, budowa, zastosowanie. *Zeszyty Naukowe AGH nr 1260, Automatyka* 47, 9–46 (1989)
26. Tadeusiewicz, R., Pachowicz, P.: CESARO - nowy system automatycznej analizy obrazów. *Wiadomości Telekomunikacyjne* 5-6, 27–28 (1984)
27. Tadeusiewicz, R.: Metody rozpoznawania obrazów i ich medyczne zastosowania. *Jubileuszowy Zjazd Towarzystwa Chirurgów Polskich, Kraków*, 90–93 (1989)
28. Tadeusiewicz, R.: Wieloprocessorowe przetwarzanie obrazów, *Szkoła Wizji Komputerowej i Sztucznej Inteligencji, Polska Akademia Nauk, IPPT, Mądralin*, 24–36 (1989)
29. Tadeusiewicz, R.: *The Multiprocessor Architectures for Image Processing*, Osolineum, Wrocław. *Lecture Notes on Computer Vision and Artificial Intelligence*, pp. 223–269 (1990)
30. Tadeusiewicz, R.: Modele sieci neuropodobnych i przetwarzania informacji w biologicznych systemach percepcyjnych. *IV Ogólnopolskie konwersatorium, Cybernetyka- inteligencja- rozwój, PTC, Siedlce*, 119–152 (1989)
31. Tadeusiewicz, R.: Inteligentne oko robota. *Sprawy Nauki* 5, 20–21 (1992)
32. Tadeusiewicz, R.: Systemy wizyjne robotów przemysłowych. In: *WNT, Warszawa* (1992)
33. Tadeusiewicz, R., Flasiński, M.: *Rozpoznawanie obrazów*. PWN, Warszawa (1991)
34. Tadeusiewicz, R.: Sieci neuronowe w rozpoznawaniu obrazów. In: *Mat. konf. Uniwersalność cybernetyki, Kraków*, pp. 17–18 (1993)
35. Mikrut, Z., Tadeusiewicz, R.: Metodyka eksperymentów z sieciami neuronowymi rozpoznającymi obrazy. *Zeszyty Naukowe AGH, Automatyka* 66, 7–30 (1993)

36. Tadeusiewicz, R.: Komputerowe systemy przetwarzania obrazów. Rozdział w pracy zbiorowej *Nowoczesna Technika w Kulturze, Nauce i Oświacie - komputery, audio, video, TVSat, Multimedia, Infostrady*, WOK Tarnów, 281–298 (1995)
37. Tadeusiewicz, R., Mikrut, Z.: Pattern Recognition Using Neural Networks. In: *Proc. of the 8-th International Conference System-Modeling-Control, Łódź. Artificial Neural Networks and their Applications*, vol. 3, pp. 136–144 (1995)
38. Tadeusiewicz, R.: Sieci neuronowe w przetwarzaniu i rozpoznawaniu obrazów. *Seminarium Komputerowa analiza obrazu w technologii żywności*, ART, Olsztyn, 4–33 (1994)
39. Tadeusiewicz, R.: Finding of Optimal Structure of the Neural Network for Image Processing and Pattern Recognition. *Invited Session on Neural Networks for Signal Processing. IEEE Signal Processing Symposium*, 39–46 (1996)
40. Tadeusiewicz, R.: Optyczne czytniki (I). *Magazyn Komputerowy ENTER* 6, 22–24 (1992)
41. Tadeusiewicz, R.: Optyczne czytniki (II). *Magazyn Komputerowy ENTER* 7, 22–23 (1992)
42. Mikrut, Z.: Estimation of the hidden layer size based on analysis of neural networks for handwritten digit recognition. *Appl. Math. and Comp. Sci.* 4(3), 343–352 (1994)
43. Gorgoń, M., Mikrut, Z., Tadeusiewicz, R.: Image Processing Systems at Biocybernetic Lab of AGH, from TTL to Modern FPGA. In: *Colnaric, M., Adamski, M., Węgrzyn, M. (eds.) Real-Time Programming*, pp. 189–194. *IFAC Publications*, Elsevier (2003)
44. Gorgoń, M., Wiatr, K., Mikrut, Z., Tadeusiewicz, R.: Rekonfigurowalne architektury systemów sprzętowych do przetwarzania i analizy obrazów w pracach AGH. *Elektronizacja - podzespoły i zastosowania elektroniki 7-8*, 29–33 (2003)
45. Tadeusiewicz, R., Gorgoń, M., Wiatr, K., Mikrut, Z.: Reconfigurable Image Processing Architectures - Research and Current State of Art at the AGH Technical University. In: *Plaks, T.P., Athanas, P.M. (eds.) Proceedings of the International Conference on Engineering of Reconfigurable Systems and Algorithms ERSA 2002, Las Vegas*, pp. 160–166 (2002)
46. Gorgoń, M., Tadeusiewicz, R.: Hardware Based Image Processing Library for Virtex FPGA. In: *Schewel, J., Athans, P.M., Dick, C.H., McHenry, J.T. (eds.) Reconfigurable Technology - FPGAs for Computing and Applications II, Proceedings of SPIE, Boston*, vol. 4212, pp. 1–10 (2000)
47. Tadeusiewicz, R., Korohoda, P.: Komputerowa analiza i przetwarzanie obrazów. In: *Wydawnictwo Fundacji Postępu Telekomunikacji, Kraków* (1997)

VICAL: Visual Cognitive Architecture for Concepts Learning to Understanding Semantic Image Content

Yamina Mohamed Ben Ali

Research Group on Artificial Intelligence, Department of computer Science,
Badji Mokhtar University BP 12, Annaba, Algeria
e-mail: benaliyam2@yahoo.fr

Summary. In this paper, we are interested by the different sides of the visual learning and the visual machine learning, as well as the development of the "visual cognitive" evolution cycle. For this purpose, we present an expected cognitive architecture framework to highlight all the visual learning functionalities. Despite the fact that our investigations were based on the conception of a cognitive processor as a high interpreter of object recognition tasks, we strongly emphasize on a novel evolutionary pyramidal learning. Indeed, this elaborated learning approach based on association rules enables to learn highest concepts induced from concepts of lower level in order to progressively understand the highest semantic content of an input image.

1 Introduction

Visual learning is an attractive trend in object recognition investigation because it seems to be the only way to build vision systems with the ability to understand a broad class of images. Indeed, visual learning is a complex task that usually requires problem decomposition, big amount of data processing, and eventually an expensive time consuming. In most approaches to visual learning reported in the literature, learning is limited to parameter optimization that usually concerns a particular processing step, such an image segmentation, feature extraction, etc. Only, a few contribution attempt to close the feedback of the learning process at the highest (i.e. recognition) level [4, 12, 13, 14, 15, 16, 17, 19, 21].

Acknowledging the need for providing image analysis at semantic level, research efforts set focus on the automatic extraction of image descriptions matching human perceptions. The ultimate goal characterizing such efforts is to bridge the so called semantic gap between low-level visual features that can be automatically extracted from the visual content and the high-level concepts capturing the conveyed meaning. In the followed cognitive architecture we put emphasize on the fact that several levels of description are based on

the widely held belief that computational vision cannot proceed in one single step from signal-domain information to spatial and semantic understanding.

In this paper, we are interested by the different sides of the visual learning and the visual machine learning, as well as the development of the "visual cognitive" evolution cycle. For this purpose, we have focused our investigations on (i) a novel cognitive architecture able to outperform all visual tasks in the domain of object recognition, and (ii) the approach to interpret from object recognition the higher image understanding.

2 Visual Cognitive Framework VICAL

Though proven the layered VICAL conceptualization from input image to its understanding, we propose a visual machine learning that implements: a *cognitive* behavior for *visual modality*, and a *visual learning* for object recognition. In below, we point out all fundamental features, structural and behavioral, needed in the image preprocessing (features detection) and post-processing (image understanding content-based).

2.1 Eye Processor

With respect to human sensorial organ, eye processor designs the input sensory channel which filters visual problems with the ability of perceiving what must be effectively done relating to the given problem. Moreover, in the elaborated visual architecture, eye processor represents the interface which connects the environment surrounding the *problem space* to the target cognitive behavior.

Its primary functions are limited to analyzing, processing, and interpreting visual information using some elementary operators that refer to low-level image processing. Thus, provided that visual information is somewhere in the image, eye processor encapsulates more than one elementary processor (blind detector processors) to in carry out the segmentation process in a parallel way. This strategy enables an effective exploration without any priori knowledge and reduces the segmentation time consuming.

- (i) At coarse processing granularity, the planning schedule of eye processor consists in the following activities:
 - Selection of a visual problem from problems space.
 - Decomposition of the "visual problem space" in more then one visual constraint (object) in order to extract from each visual sub-space the features vector.
- (ii) At fine processing granularity, a blind detector processor elaborates a contour detection, extracts relevant features, and then constitutes the geometric model relative to the corresponding visual constraint. The blind detector is therefore capable to:

- Performs an arithmetic mental process (reasoning process) using a buffer memory as a working memory to store, at each detecting step, relevant information needed to encode the geometric model at the next step.
- Returns the corresponding geometric model G_i as a symbolic vector of the smallest object O_i having significance with human comprehension.

Fig. 1 clearly shows that there is no interaction between blind detectors what explains the absence of a communication scheme. This is substantially correct if we consider that detector processors perform independently blind explorations with regard to the whole visual context. Relating to the fixed

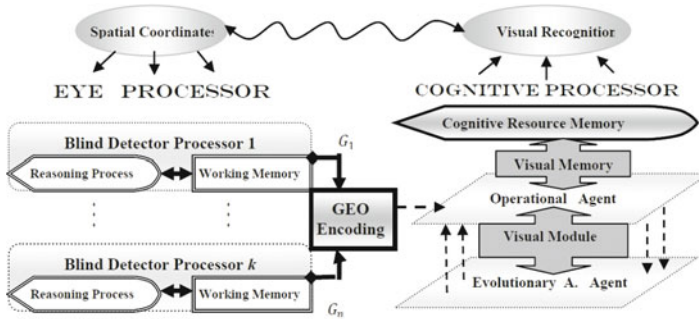


Fig. 1. Cognitive visual memory architecture

aim, the description of the geometric is subjective, because our investigations are concentrated at the cognitive level in which the bridge between the low-level processing and the high-level processing fills the semantic gap. A geometric model can be regarded as a vector of interest points of an object in image processing.

2.2 Cognitive Processor

Cognitive processor is a vulgar imitation of the brain in solving problem since logical, complex, and high functionalities such; deduction recognition and understanding are borrowed from the brain functionalities. The cognitive processor is the core of VICAL that lead to desired goal as a result of the interaction occurring at different levels of organization and at different time scales and involving not only the elementary elements (i.e. cognitive resource memory, visual module) but also the behavioral properties emerging from the interactions. Among these properties, we can cite:

- Adaptive behavior, where each input image requires its own resolution scheme improved by the cognitive processor

- Decentralized organization guided by the role of each component.
- Distributed processing with both communicative behavior and cooperative behavior between concerned components (agents).

Cognitive resource memory

Memory is one of the primary domains examined by cognitive psychologists, since encoding, storage, and retrieval of information constitute a significant portion of our cognitive activity. In the context of computer vision, and based on the fact that the memory is context-independent, we define the dominant source of knowledge as a cognitive resource memory.

Definition 1 (*Cognitive resource memory*): *The cognitive resource memory is a resource repository of multiple connected memories areas whereas each area is roughly expected for a cognitive modality (e.g. visual modality).*

This definition explains that the cognitive resource memory is not presented like a compact whole but in the form of resources network composed by several semantic nodes, each node models one modality of the cognitive memory. Indeed, the cognitive processor as a supervisor disposes of many sensorial agents (captors) to proceed with the environment resources like images, and speech. The acquisition of knowledge is done through cognitive channels under a selective activation. The distribution of knowledge implies that multiple recognition process could be undertaken in parallel and in a collaborative way. Since the role of each area memory is to enclose a priori knowledge and posteriori conclusions the cognitive processor disposes of specialized cognitive modules to reason about distinct cognitive activities.

In VICAL, we only refer to the visual modality as the effective behavior without worrying about other interesting emergent behaviors. This means a real interest only to basic components required by the visual application.

Visual memory

It is the more concerned area belonging in the cognitive resource memory. We define the visual memory as the visual information repository activated by its corresponding selective channel that points out one of the multi cognitive modalities assigned to the cognitive processor. The basic elements stored and retrieved from visual memory are concepts.

Definition 2 (*Concept*): *A concept is a symbolic formalization of concrete information in the image possessing semantic content. A concept can be either elementary like: eye, nose, arm, square, etc. or composed like: face, human, cat, house, etc. So, all knowledge acquired through eye processor enriches both visual memory, because of the direct connection established between the two reliable components, and others memory areas in case where the needed information can be useful in the search process. Thus, two meaningful searching strategies are viewed:*

- *Intra search memory.* Retrieval process of concepts consists in local interactions between some knowledge domains representing features (concepts) according to appropriate representation formalism.
- *Extra search memory.* Retrieval process of concepts consists in consulting one or more sensorial memories to locate the appropriate database representing the claimed concepts. This operation needs to create semantic interactions as communication between different cognitive resources.

Visual Module

Before considering objectives and components of the visual module, let us briefly see again the claimed goal: *from an input image, recognize all the salient objects in order to establish the concepts mapping involved to accurately provide the high meaning to the content-image.* At first sight this seems easy to accomplish but actually it is much deeper than that appear. For this reason, we propose to capture the meaning of the image toward a novel learning approach named pyramidal learning. The understanding philosophy behind pyramidal learning is to point out the top of the pyramid in order to achieve not only the highest concept assigned to the image but also the semantic content able to furthermore indexing this image.

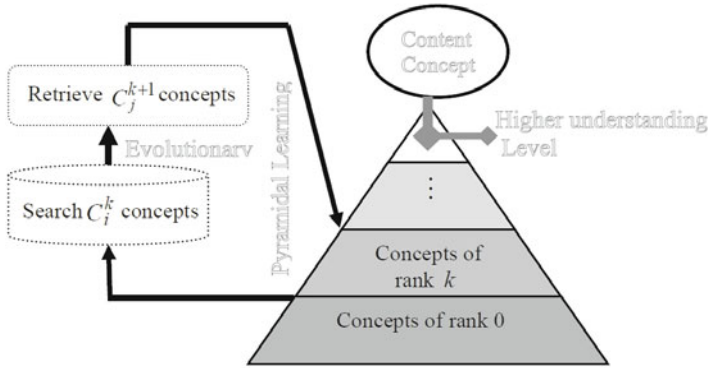


Fig. 2. Evolutionary pyramidal learning cycle

The following algorithm presents broadly the main activities elaborated by the visual module.

1. Select initially formal concepts of rank k ($k = 0$)
2. Locate the appropriate database (used for learning concepts) according to the similarity measure between the given concepts and the keywords used for indexing this database.

3. In the case where such matching is possible, thus apply the evolutionary procedure to learn high concepts and go to step 4. Otherwise, the final formal concepts are proposed as the semantic content of the proceeded image.
4. Replace the formal concepts of rank k by the new learned concepts of rank $k + 1$.
5. Compute distances between new concepts.
6. Compare these distances with the empirical threshold to ensure the fusion of some concepts and constitute the new sets of concepts.
7. Go to step 2.

Fig. 2 explains in more detail the pyramid of learning. The basis of the pyramid defines the first learning level with concepts of rank '0'. These concepts (provided by the eye processor) are obtained after decoding geometric models of objects such: $\{G_1, G_2, \dots, G_n\} \rightarrow \{C_1^0, C_2^0, \dots, C_n^0\}$. After a learning step, the concepts of superior rank ($k = 1$) are retrieved and construct the next bricks of the pyramid. This process is reiterated until the top of the pyramid is achieved by given the highest concept attributed to the image.

The above description concerns only the behavioral aspect of the visual module, remains now to describe its organizational aspect. Thus, Fig. 3 outlines the eventual multi-agent description where agents shared the same visual memory. The main activities clustered under two distinct sub-goals give more details on the specialization of the agents and their thoughtful occurrence in the cognitive processor to tackle visual recognition problems.

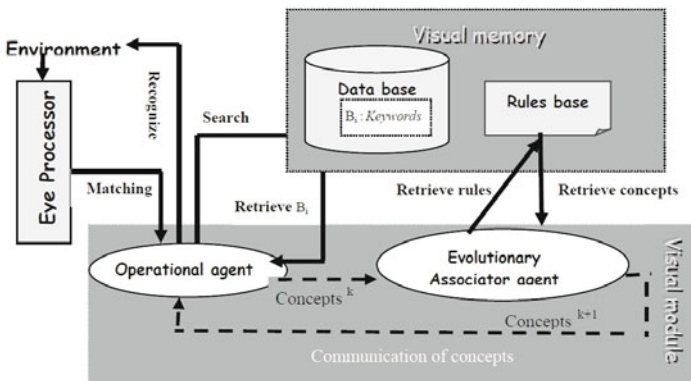


Fig. 3. Diagram of interactions between agents for solving visual problem

3 Structural Abstraction of VICAL

This section discusses the adequate organizational structure of VICAL. Effectively, based upon the fact that the recognition process is strongly influenced by two cognitive properties: evolutionary concept learning and restructuring,

the suitable organization is therefore a *role-guided* structure. A comprehensible solution that makes this possible is to settle on a *multi-agent* organization. This choice is not trivial since raise a distributed processing guided by two distinguish roles. For more comprehension, we propose the followings definitions.

Definition 3 (*Agent*): *An agent is a cognitive entity active by its deterministic role modeled by the cognitive property.*

Definition 4 (*Role*): *A role is a cognitive property that encloses some related activities.*

In agreement with all considerations, an organizational structure with two agents was being approved. The cooperation between the *operational agent* and the *evolutionary agent* through the communication of concepts enables the good progression of the iterative pyramidal learning. Thus, at each macro¹ evolution, the operational agent intend to select the target database with the predicting concepts in order to facilitate the work of the evolutionary agent, which needs the training examples to extract the relevant set of rules required to predict the target higher concept.

3.1 Operational Agent

Although decoding salient objects represents initial performed task, the operational agent pays more attention at restructuring, at each time, available data (all C_i^{k+1}) provided by the evolutionary agent. Thus, operational agent accentuates its interest on the geometric aspect of a concept rather than its semantic aspect. Thus, it manipulates objects regarded as vectors of integer coordinates, and carries out its restructuring according to the position of objects in the image.

Broadly, the cognitive activity of restructuring encloses two complementary tasks: reduction and clustering. The *reduction task* is the setting in correspondence of the nearest objects which enables the reduction of the visual space. Thus, instead of considering detected objects as spatially independent, *clustering task* aims at regrouping objects in clusters according to a vicinity relation. Such relation is considered as the inclusion (only at beginning) and becomes a relation of nearness throughout the recognition process.

Reduction task

The reduction principle is to put away the nearest objects those which are either dependent from the global context or those which belong to a local context. The reduction task enables to reduce the visual space as well as the understanding limited to a local context. Thus, we define a local context

¹ A macro generation comes after learning a concept, and then takes several micro generations needed for one evolutionary learning.

as a reduced area which provides a sense to a cluster of objects. Formally, the reduction task consists at covering nearest objects by the construction of a binary matrix $M(n \times n)$ where n represents the number of all available objects at this learning phase. The rows and columns of the matrix are then the objects $O_i (i = 1 \dots n)$ surrounded by their shapes boundaries such that $O_i = \{(Row_{min}, Row_{max}), (Col_{min}, Col_{max})\}$, and the elements are either '0', in the case where no correspondence exists between two objects, or '1' otherwise. The aimed correspondence consists at finding a logical relation between two objects of the form inclusion or neighborhood relation.

- **Inclusion**

Let us assume that the number of salient objects detected in the image (before we launch the learning algorithm) is m and the set of objects is $O = \{o_1, o_2, \dots, o_m\}$. The operational agent tries then to fill in the matrix M according to the following properties:

- R is Reflexive $\begin{cases} O_i \subseteq O_j & \text{Then } \{(i = j) \wedge M(i, j) = 1\} \\ O_i \sqsubset O_j & \text{Then } (i \neq j) \end{cases}$
- R is Asymmetric $\begin{cases} O_i \subset O_j \\ O_i \in O_j \end{cases}$
- R is Transitive $\begin{cases} (O_i \sqsubset O_j) \wedge (O_j \subset O_k) \\ (O_i \subset O_k) \end{cases}$

As is depicted in Fig. 4, the correspondence between the image and its relative matrix provides the creation of three clusters.

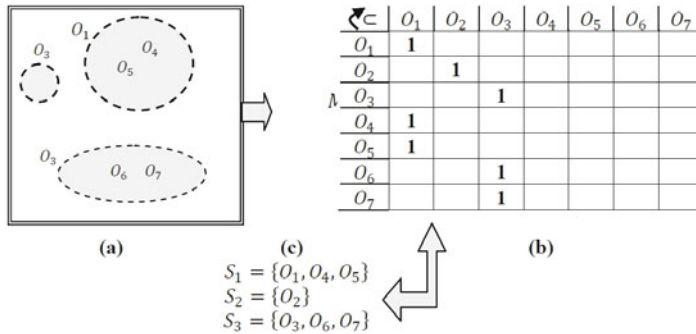


Fig. 4. Cognitive visual memory architecture

- **Neighborhood**

Let the number of reduced objects detected in the image (after at least one learning phase) is p and the set of objects is $O = \{o_1^k, o_2^k, \dots, o_p^k\}$. The index k indicates that the objects are extracted after k^{th} learning phase.

The operational agent tries then to fill in the matrix M according to the measure of the distance between pairs of objects, where in general this distance has the following properties:

$$d(O_i^k, O_j^k) = 0$$

$$d(O_i^k, O_j^k) = d(O_j^k, O_i^k)$$

Since the objects represent edges or contours we then assume that the appropriate distance is the Hausdorff distance which computes the maximum distance of a set to the nearest point in the other set [10]. More formally, Hausdorff distance between two edge points sets O_i and O_j is the maximum function (Equation 1), defined as:

$$H(O_i^k, O_j^k) = \max_{a \in O_i^k} \left\{ \min_{b \in O_j^k} \{d(a, b)\} \right\} \quad (1)$$

Thus, the operational agent assigns to the matrix elements different values according to Equation 2 as follows:

$$M(i, j) = \begin{cases} 1 & \text{If } H(O_i^k, O_j^k) \leq \omega \\ 0 & \text{If } H(O_i^k, O_j^k) > \omega \end{cases} \quad (2)$$

The parameter ω is a fixed threshold. Once M constructed by filling in its elements, we can then pass to the next step of clustering.

Clustering task

The clustering principle consists at regrouping in a cluster all the nearest objects. This step strongly depends on the reduction step and effectively enables the extraction of all clusters available at this pyramidal learning phase. The clustering task is then lunched after the construction of M . According to a column k , the existence at least of an element with value 1 at a row l yields the construction of a new cluster j . For this same column k , if there exists more than elements with values 1, then all other objects representing the rows are included in the this same cluster j (Equation 3). In the case, where all the elements of a column are set to zero this implies that the object of this column is already contained in a previous cluster.

$$\text{If } M(k, l) = 1 \text{ Then } \{O_k \in S_j \wedge O_l \in S_j\} \quad (3)$$

When all matrix columns were visited, therefore the optimal number of clusters is determined. The following algorithm enables the formation of clusters:

The result provided from this task is the number of clusters constructed as well as the covered objects contained in each cluster. It is important to claim that each cluster S_j possesses a *reference object* S_j^{ref} which represents the identifier of both the cluster and this local context. The reference object

Algorithm 1. Clustering Algorithm.

```

// In [M:matrix ( $n \times n$ )]
begin
 $j \leftarrow 0; k \leftarrow;$  //  $k$  is the column index
while  $k \leq n$  do
  if ( $(k > 1)$  and  $O_k \in \{S_e/r = 1, \dots, j - 1\}$ ) then
     $k \leftarrow k + 1$ 
  else
     $j \leftarrow j + 1; S_j = \emptyset;$ 
    for each row  $l$  to  $n$  do
      if  $M(k, l) = 1$  then
         $S_j = S_j \cup O_l;$ 
        if  $(k = l)$  then
           $S_j^{ref} = \{O_l\};$ 
        end if
      end if
    end for
  end if
end while
// Out [S: set]
end.
```

is useful to compute the different distances between different clusters (local contexts). In the case of inclusion relation, S_j^{ref} is only one object that which matches to the encompassing object. In the case of a neighborhood relation, the reference object will be the union of all objects present in the current cluster.

3.2 Evolutionary Associator Agent

The main characteristic of learning-based approaches is their ability to adjust their internal structure according to input and respective desired output data pairs in order to approximate the relations (rules) implicit in the provided training data, thus elegantly simulating a reasoning process. Consequently, the use of some approaches like classification rules or even association rules provides a powerful method for discovering complex and hidden relationships for a variety of applications domains.

Association rules

Association rules are used at a high interpretation and representation granularity to emerge strongly associated objects in order to get a more compact representation of the data. We define association rules [2] according to the context of perceptual objects. Let $O = \{c1, c2, \dots, cn\}$ be a set of objects concepts, and D be a database (a set of formal concepts hierarchies), where each formal concept hierarchy h is a set of objects such that $h \subseteq O$. We say

that a formal object hierarchy h contains X , a set of objects in O , if $X \subset h$. An association rule is an implication of the form $X \Rightarrow Y, X \subset O, Y \subset O$, and $X \cap Y = \emptyset$.

Many important algorithms were used to discover association rules [3, 18]. This formulation yields a thought of two basic aspects for data representation:

- How to discover the appropriate association rules from a set of concepts X detecting some relationship between the predicting concepts and the goal concept Y . Actually, with regard to all consideration cited above, the problem is tackled by considering it as a concept learning problem. In this case, each association rule is subjected to learn one high formal concept.
- If we consider the vector space whose dimensions are the concepts objects, concept hierarchy can be represented as object vectors with binary coordinates denoting the occurrence of objects in the hierarchy.

Visual evolutionary-based learning

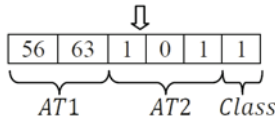
Representing concepts as sets of rules has long been popular in machine learning, because, among other properties, rules are easy to represent and humans can interpret them easily. In evolutionary algorithms there are two main ways to represent rule sets. In the "Michigan" approach exemplified by Holland classifier [9], and the "Pittsburgh" approach exemplified by Smith LS-1 system [5, 20]. Systems using the Michigan approach maintain a population of *individual rules* that compete with each other for space and priority in the population. This approach is simpler and syntactically shorter. This tends to reduce the time taken to compute the fitness function and to simplify the design of the genetic operators. A number of systems based on concept learning with Michigan approach have been proposed COGIN [6] and REGAL [7].

In contrast, systems using Pittsburgh approach maintain a population of variablelength rule sets that compete with each other with respect to performance on the domain task. This approach takes into account rule interaction when computing the fitness function of an individual. Consequently, some systems based on concept learning with Pittsburgh approach have been proposed like GABIL [5] GIL [11] and DOGMA [8].

Individual (rule) representation

In our studies, we adopt the Michigan approach to encode a rule. Thus, an individual in its general form is a set of items of $attribute_i \in [l_i, u_i]$, where $attribute_i$ is the i^{th} numeric attribute in the rule template from the left to the right. Fig. 5 outlines a general example:

This rule specification could induce to some problems in recombination process. Indeed, if two individuals have different representations the crossover applied to them will generate invalid offspring. Hence, to avoid to have invalid offspring due to the absence of some attributes in a rule, and in order to

*Do**Ru***Fig. 5.** Hybrid encoding of individual

use the usual crossover without specific modifications, each attribute in the individual has the same position (rank) that it occur in the dataset (in the database).

Evolutionary learning process

With respect to the used dataset where all attributes are nominal, the absence of an attribute in the dataset is referred to a gene with '0' value. The evolutionary learning process is performed by the function *EvoAlg* as shown in Fig. 6. Thus, *EvoAlg* has a set of examples as its input parameter. It returns a rule that is the best individual of the last generation. The initial population P is built randomly by the function *InitializePopulation*. Some examples are randomly selected and individuals that cover such examples are generated. After initializing the population, the for-loop repeats the evolutionary process *max-generations*. At each iteration, the individuals of the population are evaluated according to defined fitness function, thus each individual acquires goodness. The best individual of each generation is replicated (*Replicate*) to be included in the next generation. Later, a set of individuals are selected through the roulette wheel method and replicated. Finally, another set of individuals are recombined and the offspring are included in the next generation.

```

Function EvoAlg
// In [T: file of encoded-examples ]
Begin
InitializePopulation(P)
For i=1 to max-generations do
Evaluate(P);
next.P:=SelectTheBestOf(P);
next.P:=next.P+Replicate(P);
next.P:=next.P+Recombine(P);
P:=next.P
End;
Evaluate(P);
return SelectTheBestOf(P) // Out [ r:Rules]
End EvoAlg

```

Fig. 6. Pseudocode of evolutionary learning rules

Equation 3 gives the fitness function $f(\tau)$ used during the evaluation process. The greater the value, the better the individual is.

$$f(\tau) = N - CE(\tau) + G(\tau) + coverage(\tau) \tag{4}$$

Where τ is an individual, N is the number of examples being proceeded; CE is the concept error, i.e. the number of examples belonging to the region defined by the rule τ , which they do not have the same concept (class); $G(\tau)$ is the number of examples correctly "classified" by τ , and $coverage(\tau)$ gives the size proportion correctly "classified". For more detail about the influence of $coverage(\tau)$ on fitness see [1].

4 Experimental Results

To test our image theory we created synthetic images as a starting point in showing the feasibility of concept learning as a technique to understanding content image. The images database contains combinations of plane geometric shapes, where each shapes belongs to the set $S = \{ \text{triangle, circle, rectangle, hexagon, ellipse} \}$. The number of images in the database generated for the test phase corresponds to 30 images, where each image is identified by a label. As depicted in Figure 7, the majority of images have only one understanding level. This constraint is necessary because of the prediction of concepts. But to be able to validate at least two learning levels, we have drawn four images (20, 24, 27, 30) which represent semantic objects, constructed from S (primitives concepts), and correspond respectively to face, bicycle, chair, and cat. The association rules shown to the environment have the form:

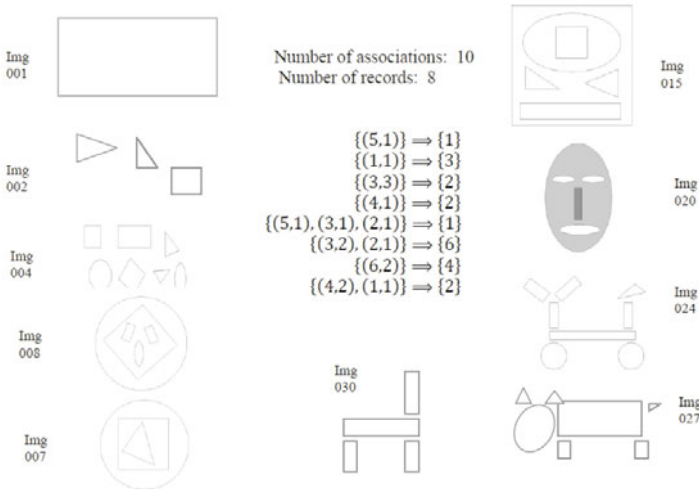


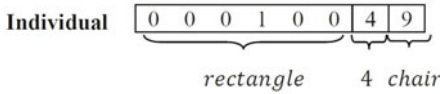
Fig. 7. Geometric shapes and association rules as selected training examples

$$\{(AT1_1, AT2), (AT1_2, AT2), \dots, (AT1_k, AT2), \} \Rightarrow \{AT1_{k+1}\}$$

In concordance with the chosen images, we define the new application context:

$$Domain \begin{cases} AT1_i \in \{\text{triangle, circle, rectangle, hexagon, ellipse}\} \\ AT2 \in [1, 10] \end{cases}$$

Example of coding rule. Rule : If $AT1_1 \in \{\text{rectangle}\}$ and $AT2 \in [1, 5]$ Then chair



Each image in the database stores both the low-level features (first descriptor) and the high-level concept attributes (second descriptor). In fact, at this postprocessing level, we are only interested by the second descriptor. Thus, the first rule $\{(5, 1)\} \Rightarrow \{1\}$ tells us that if there is a hexagon there is triangle. This result is validated according to a high fitness attributed to this rule. Other rules indicate the same reasoning such that the rule $\{(1, 1)\} \Rightarrow \{3\}$ where if there is a triangle there is a square. Also, the rule $\{(4, 1)\} \Rightarrow \{2\}$ illustrates that if there is a rectangle there is necessary a circle. Furthermore, to show the ability of the system to learn more complex objects not composed only by primitive concepts, we have introduced four specialized rules. Each rule yields to each corresponding concept (face:7, bicycle:8, chair:9, and cat:10).

$$R1 : \{(6, 4), (4, 1)\} \Rightarrow \{7\}; \quad R2 : \{(4, 5), (1, 1), (2, 2)\} \Rightarrow \{8\}$$

$$R3 : \{(4, 4)\} \Rightarrow \{9\}; \quad R4 : \{(4, 3), (1, 3), (6, 1)\} \Rightarrow \{10\}$$

In the current studies, the number of occurrence of each object must appear in order to avoid the redundancy of the same objects. identifiers as conditions in the left side of a rule. We have deliberately omitted to reserve a field of distance in the object representation since the coverage of objects in a cluster is a task performed before learning rules. Then, we suppose that all nearest objects could as a whole specify a new concept. Obviously, we want to stress that our synthetic images are not as complex as images from the real world. But as a first attempt to mine association rules with evolutionary learning in image processing seems a good trend to improve the emergent behaviour from the cognitive architecture.

5 Conclusion

We present in this paper a cognitive architecture inspired from human characteristics to deal with the visual cognitive modality. The system includes many reasoning paradigms such as image processing, inductive learning, evolutionary computing and image mining. Our efforts were based on the conceptualization of the cognitive framework for detecting and understanding objects in

several contexts. Obviously, the system relies on knowledge discovering using association rules from databases. Results obtained so far look promising but we need to improve several aspects in our research efforts. We can currently working on the some evident tasks like image mining, and image retrieval to outperform the semantic content-based image. These visual tasks will help in the automation of such behaviours. Also, we should give more attention to other kind of attributes (discrete and continuous) and how to determine the optimal representation and even the appropriate genetic operators. In the future, we look to improve the cognitive aspect of the architecture by introducing more than one cognitive modality, and to find the optimal way to connect the aimed specialized components to this general cognitive framework. Regarded to the fixed objective, it results therefore a good opportunity to develop both intelligent and autonomous system in the domain of pattern recognition.

References

1. Aguilar-Ruiz, J.S., Riquelme, J.C., Toro, M.: Evolutionary learning of hierarchical decision rules. *IEEE Transactions on Man and Cybernetics, Part B* 33(2), 324–331 (2003)
2. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between set of items in large databases. In: *Proc. ACM SIGMOD International Conference on Management of Data*, Washington, DC, pp. 207–216 (1993)
3. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large database. In: *Proceedings of 20th International Conference on Very Large Data Bases*, Santiago, Chile (1994)
4. Bhanu, B., Lin, Y., Krawiek, K.: *Evolutionary synthesis of pattern recognition systems*. Springer, New York (2005)
5. De Jong, K.A., Spears, W.M., Gordon, D.F.: Using genetic algorithms for concept learning. *Machine Learning*, 161–188 (1993)
6. Giordana, A., Neri, F.: Search intensive concept induction. *Evolutionary Computation* 3(4), 376–416 (1996)
7. Greene, D.P., Smith, S.F.: Competition-based induction of decision models from examples. *Machine Learning* 13, 229–257 (1993)
8. Hekanaho, J.: GA-based rule enhancement concept learning. In: *Proc. Third International Conference on Knowledge Discovery and Data Mining*, New Port Beach, CA, USA, pp. 183–186 (1997)
9. Holland, J.: Escaping brittleness: The possibilities of general-purpose learning algorithms allied to parallel rule-based systems. In: Michalski, R., Carbonell, J., Mitchell, T. (eds.) *Machine Learning: An Artificial Intelligence Approach* (1986)
10. Huttenlocher, D.P., Rucklidge, W.J.: A multi-resolution technique for comparing images using Hausdorff distance, Dpt. of CS, Cornell University (1994)
11. Janikow, C.Z.: A knowledge intensive genetic algorithm for supervised learning. *Machine Learning* 13, 189–228 (1993)

12. Krawiec, K.: Learning high-level visual concepts using attributed primitives and genetics programming. In: Rothlauf, F., Branke, J., Cagnoni, S., Costa, E., Cotta, C., Drechsler, R., Lutton, E., Machado, P., Moore, J.H., Romero, J., Smith, G.D., Squillero, G., Takagi, H. (eds.) *EvoWorkshops 2006*. LNCS, vol. 3907, pp. 515–519. Springer, Heidelberg (2006)
13. Krawiec, K.: Evolutionary learning of primitive-based visual concepts. In: *Proc. IEEE Congress on Evolutionary Computation*, Vancouver, BC, Canada, pp. 1308–1315 (2006)
14. Krawiec, K., Bhanu, B.: Visual learning by coevolutionary features synthesis. *IEEE Transactions on Systems Man and Cybernetics Part B* 35, 409–425 (2005)
15. Maloof, M., Langley, P., Binford, P., Nevatia, R., Sage, S.: Improved rooftop detection in aerial images with machine learning. *Machine Learning* 53, 157–1991 (2003)
16. Ogiela, M., Tadeusiewicz, R.: Nonlinear processing and semantic content analysis in medical imaging - a cognitive approach. *IEEE Transactions on Instrumentation and Measurements* 54, 2149–2155 (2005)
17. Rizki, M., Zmuda, M., Tamburino, L.: Evolving pattern recognition systems. *IEEE transactions on Evolutionary Computation* 6, 594–609 (2002)
18. Savasere, A., Omiecinski, E., Navathe, S.: An efficient algorithm for mining association rules. In: *Proceedings of the VLDB Conference*, Zurich, Switzerland, pp. 432–444 (1995)
19. Segen, J.: GEST: a learning computer vision system that recognizes hand gestures. In: Michalski, R., Tecuci, G. (eds.) *Machine Learning, A Multistrategy Approach*, vol. IV, pp. 621–634. Morgan Kaufman, San Francisco (1994)
20. Smith, S.: Flexible learning of problem solving heuristics through adaptive search. In: *Proc. Eight International Joint Conference on Artificial Intelligence*, pp. 422–425 (1983)
21. Teller, A., Veloso, M.: PADO: a new learning architecture for object recognition. In: Ikeuchi, K., Veloso, M. (eds.) *Symbolic Visual Learning*, pp. 77–112. Oxford Press, New York (1997)

Implementation of Computer Vision Algorithms in DirectShow Technology

Piotr Szczypiński, Paweł Pełczyński,
Dominik Szajerman, and Paweł Strumiłło

Institute of Electronics, Technical University of Łódź,
Wólczańska St. 211/215,
e-mail: piotr.szczypinski@p.lodz.pl, pawel.pelczynski@p.lodz.pl
drs@ics.p.lodz.pl, pawel.strumillo@p.lodz.pl

Summary. In this paper an application of DirectShow software technology in a computer vision system is described. DirectShow (DS) is well suited for image processing and analysis tasks. In the reported study it was successfully applied in a computer stereovision system. Physical cameras of the system are represented by DS source filters connected to image analysis procedures. Original filter prototypes were designed for stereo disparity estimation and scene analysis tasks. Image analysis procedures for scene depth estimation were build and tested. The developed system has proved usefulness of the DirectShow technology in computer vision applications.

1 Introduction

Development of vision systems that employ multiple hardware and software resources is not a straightforward task. Such systems should be of modular structure and have flexible architecture. The modular design is important in projects including a group of programmers working on different aspects of the project. The system should be flexible not to depend on specific hardware or hardware versions and it should automatically adapt to various configurations. Another requirement in vision systems is a need for conversion of data forms and formats. One of such problems is object tracking in image sequences. The input is a video data whilst the output is usually a motion defined by coordinates, motion vectors and rotation angles. Moreover, multicamera vision systems such as stereovision systems require data synchronization and parallel data processing (multithreading, pipelining and graphics processing units general-purpose computing). The requirements of vision systems programming and achieving acceptable programming efficiency in such vision systems can be satisfied by employing multimedia frameworks - technologies for multimedia software development. The goal of this publication is to present the application of DirectShow multimedia framework in

a stereovision system for detection of objects, object localization, tracking and its parametric description.

DirectShow (DS) is a software technology for multimedia applications in Microsoft Windows operating systems [1], [2]. It serves as a framework for development of video and audio processing programs and modules. DS is one of the three multimedia technologies available for Windows platforms. The two other are Video for Windows and Media Foundation [3]. One of the main features of DS is its modularity. DS modules are called filters. They could be developed by different groups of programmers or software producers independently. Filters are dynamically connected into the so called graph in the time of program execution. There is no need to compile all the modules into a single application. An example of DS based video player is shown in Fig. 1. It consists of the following filters: video file reader, audio video splitter, decompression of audio and video, and presentation filters.

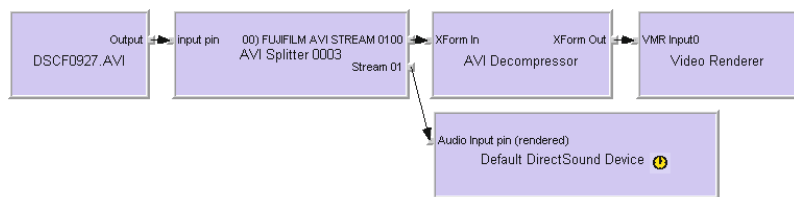


Fig. 1. Graph of filters for video file playback

Filters are implemented as DLL libraries. They are developed according to the specification of Component Object Model (COM) technology. The main assumptions are:

- each filter is a separate COM object,
- each filter is identified by its unique GUID number,
- filter properties are stored in the Windows registry file.

Connection of filters can be a fully automatic process. The type of filter used and its possible connections depend on a set of filters available in a given computer. Thus, the same multimedia file can be played by using different filters or cannot be if the important filter lacks. DS application interacts with filters via intermediary COM object called filter graph manager. It builds a graph by using IGraphBuilder. Graph is a set of connected filters. If necessary, additional transform filters are added to provide compatibility of different data stream formats between filters. The main application can also point out specific filters to be used or even specify connection order. Intermediary object is also used to control filter functions, e.g. play, rewind or pause of the video stream.

2 Filter Programming

DS filter is largely a self-processing module with a significantly greater number of functions and tasks than would result from the operation of the algorithm for processing multimedia data. This module is equipped with interfaces to communicate with other filters, intermediate objects (GraphBuilder) and the main program. These methods are public, available (can be called from outside the filter) and abstract (undefined). Filter interfaces are identified by individual, unique GUID numbers and they have to be compatible with COM technology. A typical interface is the interface of filter terminal, called IPin. Its methods are listed below.

Table 1. An example of IPin interface methods

Method	Description
Connect	Connects the pin to another pin
ReceiveConnection	Accepts a connection from another pin
Disconnect	Breaks the current pin connection
ConnectedTo	Retrieves the pin connected to this pin
ConnectionMediaType	Retrieves the media type for the current pin connection
QueryPinInfo	Retrieves information about the pin
QueryId	Retrieves the pin identifier
QueryAccept	Determines whether the pin accepts a specified media type
EnumMediaTypes	Enumerates the pin's preferred media types
QueryInternalConnections	Retrieves the pins that are connected internally to this pin
EndOfStream	Notifies the pin that no additional data is expected
BeginFlush	Begins a flush operation
EndFlush	Ends a flush operation
NewSegment	Notifies the pin that media samples are grouped as a segment
QueryDirection	Retrieves the direction of the pin (input or output)

IPin interface provides the methods:

- for negotiating data formats (EnumMediaTypes),
- for connecting and disconnecting terminals (Connect, Disconnect),
- for identification (QueryID),
- for data flow control (EndOfStream, BeginFlush, EndFlush) and others.

IPin interface inherits input and output object classes, e.g. CBaseInputPin and CBaseOutputPin. These classes provide some additional methods for multimedia data transfer. A fully functional filter provides multiple interfaces of the methods. These are the interfaces of individual pins, the interface indicating filter resources, a filter work control interface, the interface of dialog box for filter properties, etc.

The methods included in the interface must be implemented (written) by the developer creating a filter. It is necessary to provide adequate communication between the main program, an intermediary object and the neighbouring filters. This communication deals with several aspects: filter interconnection

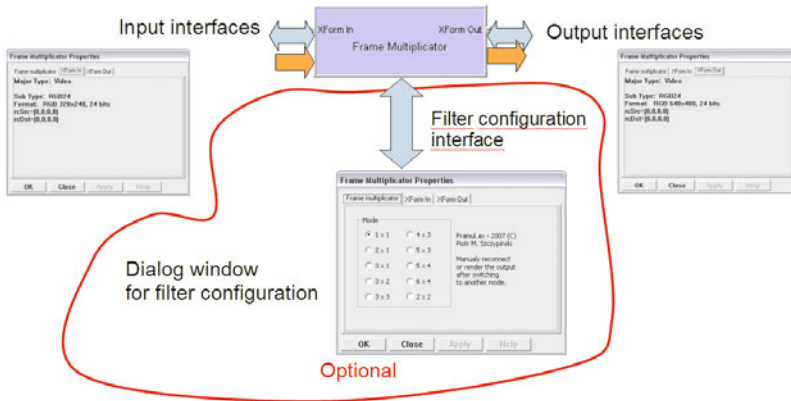


Fig. 2. Filter interfaces

strategy, negotiation of transmitted data formats, interface and pin identification, control of data stream playback, processing and synchronization, management of threads for filter control and data transfer, etc. The flexibility of the DS brings with it a significant complication of the process of creating a filter.

Writing the filter from scratch would be extremely laborious without base classes, which were made available by Microsoft. Base classes implement a part of more trivial methods, necessary for the proper functioning of the filters with specific functionality. While creating own filter one creates a new class that inherits from the selected base class. For example, writing a data-processing filter the following base classes can be used: `CTransformFilter`, `CTransformInputPin` and `CTransformOutputPin`. Creating a source filter (eg, providing data to graph from the outside) or output filter (presentation), we can use base classes: `CSource` and `CRenderedInputPin`.

Necessary Tools for creating DS filters are: the development environment (it is sufficient to use Visual C++ Express Edition [4]), Windows Platform SDK (includes base classes and tools for testing filters), the program Guidgen, the `regsvr32` tool and GraphEdit program (included in the Windows Platform SDK) or Graph Studio. Visual C++ environment is used to create filter's source code and its compilation. The environment allows also to compile a library of base classes available in the Windows Platform SDK. Guidgen program is used to generate GUID numbers identifying the filter and its interfaces. `Regsvr32` program is available in Windows system directory and is used to register a new filter in an operating system. Programs GraphEdit and GraphStudio are used for graph visualization, manual filter connection into graphs and testing their performance.

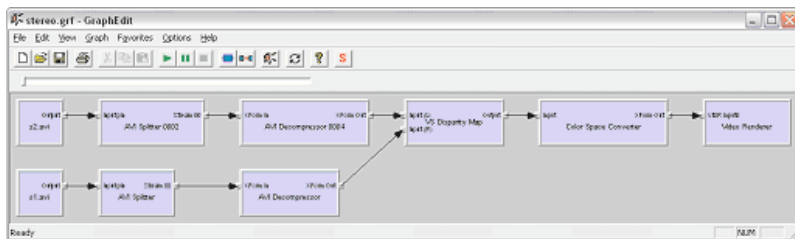


Fig. 3. Filter graph for disparity estimation in GraphEdit window

3 Filter Pattern Design

The aim of this study was to develop a methodology for the design of vision systems for scene analysis using DirectShow technology. Implementing the system using DS required to develop filters with specific functionality, number of inputs and outputs, control, types of data formats for input and output. The system includes filters for calculating disparity map, object detection and presentation of analysis outcomes. In order to integrate algorithms that were developed earlier, filters with the required functionality need to be defined in DS system. Filter source code was written in C++, using the COM and DS specification, and DS base classes. The written source code is a package of algorithms for image processing and analysis. It is fully functional in terms of DirectShow technology, but the code need to be supplemented in order to obtain the system functionality. Code fragments, where additional commands need to be entered, were marked in the source code.

All data in the DS system are transferred by objects belonging to the class inheriting from the IMediaSample interface. This interface allows to download information about the index to a block of data transmitted and the size of the block (methods: GetPointer and GetActualDataLength). Format of the received data or data sent through the filter is determined in the negotiation process. To send data about data formats CMediaType class objects are used (containing AM_MEDIA_TYPE structure) in the described project. Structure AM_MEDIA_TYPE carries information about the main type (majortype), subtype, and optionally additional information tables set out by the field Formattype. These additional tables may be e.g. VIDEOINFOHEADER for imaging data.

Developed filters operate on two types of multimedia data, namely: uncompressed image data and a table describing a result of image sequence analysis, such as the scene model parameters.

1. Imaging data: majortype: MEDIATYPE_Video, Subtype: MEDIASUBTYPE_RGB32 or MEDIASUBTYPE_RGB24. These are colour images of 32 or 24 bits per pixel. Note that DS does not define the greyscale image format.

2. The data describing the objects' parameters: majortype: MEDIATYPE_Video, Subtype: MEDIASUBTYPE_NAVIOBJ.

Filter development procedure was defined as a set of the following modifications:

1. Find the header file, which defines the filter GUID identification numbers,
2. Generate a new GUID numbers and replace them in the header file,
3. Find the structures AMOVIESETUP_FILTER and CFactoryTemplate in the source file of the filter, and replace the names of the filters and titles of the properties dialog boxes.
4. Find a piece of code marked "@@@ Complete code here ..." and write a code of the required calculation algorithm.

4 DS Based Application for Scene Depth Estimation

On the basis of the prepared filter pattern new filter was designed. It performs initial image preprocessing and disparity estimation in a pair of stereovision images. Image processing operations were implemented in a graphics processing unit (GPU), which is very efficient in vector operations. More details about applying GPU for disparity estimation are given in [5].

The developed DirectShow filter is a dynamic-link library which offers a set of the following image processing functions. They use OpenGL in order to allow setting of the parameters, passing of the input images to the GPU, control of the processing on the GPU and fetching the results from it. The filter has two input pins and one output pin (Fig. 4). The input pins are used to retrieve images from the left and right camera. The images can have any resolution and should be in RGB colour format (24 bits per pixel). The output pin transmits the result as the image of the same resolution as the input ones. The output image is also three-channel. The output disparity is stored in the first channel. The processing in the filter consists of a few passes. Every pass

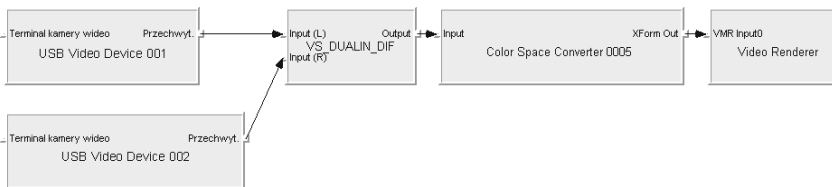


Fig. 4. Filter graph for depth estimation

is a fragment shader processing of the input textures which contain input data. The result of each pass is stored in the output texture, which is then delivered as an input to the next pass. The processing passes are shown in Fig. 5.

1. both input images are converted into greyscale and their geometry is corrected and rectified, the result is stored as a single texture ($I_{l,r}^Y$).
2. the average brightness is calculated for vertical blocks (size $1 \times b$, where b is the processing parameter - the block size).
3. the average brightness from previous pass is used to calculate the average brightness for blocks of size $b \times b$, next it is subtracted from the original brightness.
4. absolute differences between left and right image are calculated, their count depends on a processing parameter.
5. the sum of the absolute differences in the blocks (size $1 \times b$) is calculated.
6. the SAD in the blocks (size $b \times b$) is calculated.
7. the minimal from the SADs is chosen, its index is stored as the result - calculated disparity value.

The results of the described image processing procedure are shown in Fig. 6. Image disparity map is shown as a grayscale image in which closer scene

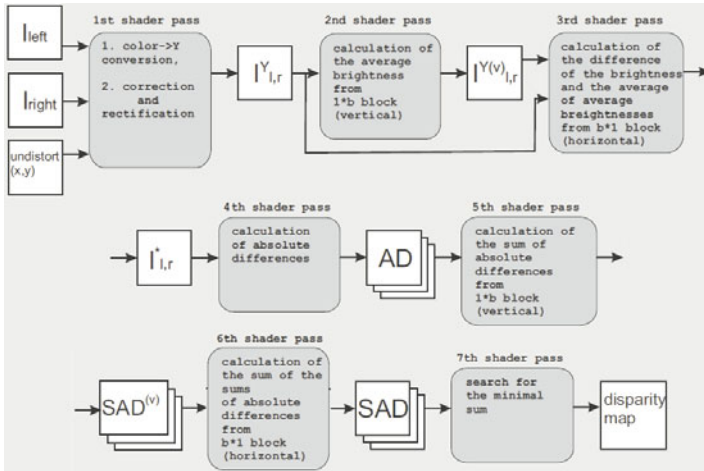


Fig. 5. Data flow in GPU processing

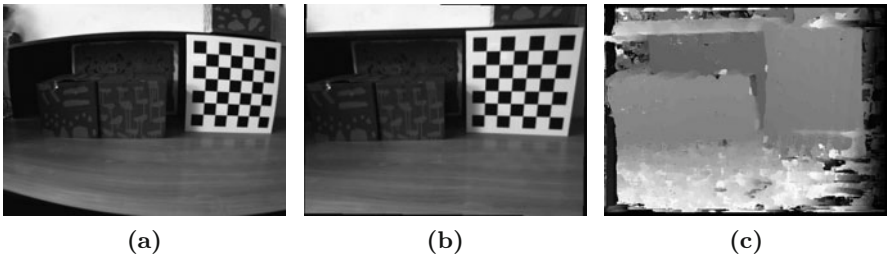


Fig. 6. Original image (a) undistorted image (b) and disparity map (c)

elements are coded by brighter map regions. GPU implementation makes execution of this procedure fast enough to be run in real time. Despite the use of GPU the developed filter is quite universal. It can process images in NVIDIA GPUs series 8000 and higher.

5 Conclusions

DirectShow software technology was used to implement computer stereo vision system. Stereo cameras were represented as DS source filters connected to the filter for image distortion correction and disparity estimation. Other filter prototypes were designed for further scene analysis procedures. A useful system for scene depth estimation was build and tested. The main problems that were solved while implementing DS multimedia framework were: synchronization of video data obtained from two cameras, definition of the original object describing data format to be transferred between modules and the implementation of advanced image processing filter in GPU platform. It was shown that advanced image processing tasks can be mapped onto DirectShow technology that offers more efficient software development and testing in a group of cooperating programmers.

References

1. Microsoft MSDN DirectShow documentation (July 7, 2007), <http://msdn.microsoft.com/en-us/library/ms783323.aspx>
2. Pesce, M.D.: Programming Microsoft DirectShow for Digital Video and Television. Microsoft Press (April 16, 2003)
3. http://en.wikipedia.org/wiki/Multimedia_framework
4. <http://www.microsoft.com/express>
5. Strumiłło, P., Szajerman, D., Pełczyński, P., Materka, A.: Implementation of Stereo Matching Algorithms on Graphics Processing Units. In: Choraś, R.S., Zabłudowski, A. (eds.) Image Processing & Communications Challenges, pp. 286–293. Academy Publishing House EXIT, Warsaw (2009)

Implementation of Hurwitz-Radon Matrices in Shape Representation

Dariusz Jakóbczak

Koszalin Technological University,
ul. Śniadeckich 2, 75-453 Koszalin
e-mail: djakob@ie.tu.koszalin.pl

Summary. Computer vision needs suitable methods of shape representation and contour reconstruction. One of them, invented by the author and called method of Hurwitz-Radon Matrices (MHR), can be used in representation and reconstruction of shapes of the objects in the plane. Proposed method is based on a family of Hurwitz-Radon (HR) matrices. The matrices are skew-symmetric and possess columns composed of orthogonal vectors. Shape is represented by the set of nodes. It is shown how to create the orthogonal and discrete OHR operator and how to use it in a process of shape representation and reconstruction. MHR method is interpolating the curve point by point without using any formula or function.

1 Introduction

An important problem in machine vision and computer vision [1] is that of appropriate shape representation and reconstruction. Classical discussion about shape representation is based on the problem: contour versus skeleton. This paper is voting for contour which forms boundary of the object. Contour of the object, represented by contour points, consists of information which allows us to describe many important features of the object as shape coefficients [2]. In the paper contour is dealing with a set of curves. Curve modeling and generation is a basic subject in many branches of industry and computer science, for example in the CAD/CAM software.

The representation of shape can have a great impact on the accuracy and effectiveness of object recognition [3]. In the literature, shape has been represented by many options including curves [4], graph-based algorithms and medial axis [5] to enable shape-based object recognition. Digital curve (open or closed) can be represented by chain code (Freeman's code). Chain code depends on selection of the started point and transformations of the object. So Freeman's code is one of the method how to describe and to find contour of the object. An analog (continuous) version of Freeman's code is the curve $\alpha - s$. Another contour representation and reconstruction is based

on Fourier coefficients calculated in Discrete Fourier Transformation (DFT). These coefficients are used to fix similarity of the contours with different sizes or directions. If we assume that contour is built from segments of a line and fragments of circles or ellipses, Hough transformation is applied to detect contour lines. Also geometrical moments of the object are used during the process of object shape representation [6]. Proposed MHR method requires to detect specific points of the object contour, for example in compression and reconstruction of monochromatic medical images [7]. Contour is also applied in shape decomposition [8]. Many branches of medicine, for example computed tomography [9], need suitable and accurate methods of contour reconstruction [10]. Also industry and manufacturing are looking for methods connected with geometry of the contour [11]. So suitable shape representation and precise reconstruction or interpolation [12] of object contour is a key factor in many applications of computer vision.

2 Contour Points Based Shape Representation

Shape can be represented by object contour, i.e. curves that create each part of the contour. One curve is described by the set of nodes $(x_i, y_i) \in \mathbf{R}^2$ (contour points) as follows in proposed method:

1. nodes (interpolation points) are settled at local extrema (maximum or minimum) of one of coordinates and at least one point between two successive local extrema;
2. each node (x_i, y_i) is monotonic in coordinates x_i or y_i (for example equidistance in one of coordinates);
3. one curve (one part of the contour) is represented by at least five contour points.

Condition 1 is done for the most appropriate description of a curve. So we have n curves C_1, C_2, \dots, C_n that build whole contour and each curve is represented by nodes according to assumptions 1-3.

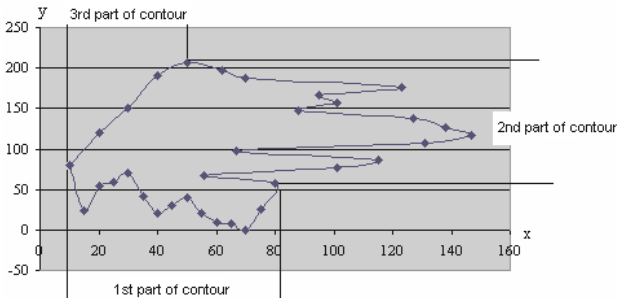


Fig. 1. Contour consists of three parts (three curves and their nodes)

Fig. 1 is an example for $n = 3$: first part of contour C_1 is represented by nodes monotonic in coordinates x_i , second part of contour C_2 is represented by nodes monotonic in coordinates y_i and third part C_3 could be represented by nodes either monotonic in coordinates x_i or monotonic in coordinates y_i . Number of curves is optional and number of nodes for each curve is optional too (but at least five nodes for one curve). Representation points are treated as interpolation nodes. How accurate can we reconstruct whole contour using representation points? The shape reconstruction is possible using novel MHR method.

3 Shape Reconstruction via MHR Method

The following question is important in mathematics and computer sciences: is it possible to find a method of curve interpolation in the plane without building the interpolation polynomials and without mathematical form of the curve? Our paper aims at giving the positive answer to this question. There exists several well established methods: spline functions [13], shape-preserving techniques [14], subdivision algorithms [15], Bezier curves, B-splines, NURBS [16] and others [12] to overcome difficulties of polynomial interpolation, but matrix interpolation MHR (based on simple matrix calculations with low computational costs) seems to be quite novel in the area of shape reconstruction. In comparison MHR method with Bézier curves, Hermite curves and B-curves (*B-splines*) or NURBS one unpleasant feature of these curves must be mentioned: small change of one characteristic point can make big change of whole reconstructed curve [17]. Such a feature does not appear in MHR method [18]. Methods of curve interpolation based on classical polynomial interpolation: Newton, Lagrange or Hermite polynomials and spline curves which are piecewise polynomials [19]. Classical methods are useless to interpolate the function that fails to be differentiable at one point, for example the absolute value function $f(x) = |x|$ at $x = 0$. If point $(0;0)$ is one of the interpolation nodes, then precise polynomial interpolation of the absolute value function is impossible. Also when the graph of interpolated function differs from the shape of polynomials considerably, for example $f(x) = 1/x$, interpolation is very hard because of existing local extrema of polynomial. Lagrange interpolation polynomial for function $f(x) = 1/x$ and nodes $(5;0.2)$, $(5/3;0.6)$, $(1;1)$, $(5/7;1.4)$, $(5/9;1.8)$ has one minimum and two roots.

We cannot forget about the Runge's phenomenon: when interpolation nodes are equidistance then high-order polynomial oscillates toward the end of the interval, for example close to -1 and 1 with function $f(x) = 1/(1 + 25x^2)$ [20] or $f(x) = 1/(1 + 5x^2)$. Method of Hurwitz - Radon Matrices (MHR), described in this paper, is free of these bad features. Complexity of calculations for one unknown point in Lagrange or Newton interpolation based on n nodes is connected with the computational cost of rank $O(n^2)$. Complexity of calculations for L unknown points in MHR interpolation based

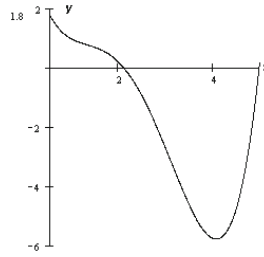


Fig. 2. Lagrange interpolation polynomial for nodes $(5;0.2)$, $(5/3;0.6)$, $(1;1)$, $(5/7;1.4)$, $(5/9;1.8)$ differs extremely from the shape of function $f(x) = 1/x$

on n nodes is connected with the computational cost of rank $O(L)$ [18]. This is very important feature of MHR method. The curve or function in MHR method is parameterized for value $\alpha \in [0; 1]$ in the range of two or more successive interpolation nodes.

3.1 The Operator of Hurwitz-Radon

Adolf Hurwitz (1859-1919) and Johann Radon (1887-1956) published the papers about specific class of matrices in 1923, working on the problem of quadratic forms. Matrices $A_i, i = 1, 2, \dots, m$ satisfying

$$A_j A_k + A_k A_j = 0, \quad A_j^2 = -I \quad \text{for } j \neq k; j, k = 1, 2, \dots, m$$

are called a *family of Hurwitz - Radon matrices*. A family of Hurwitz - Radon (HR) matrices has important features [21]: HR matrices are skew-symmetric ($A_i^T = -A_i$) and reverse matrices are easy to find $(A_i)^{-1} = -A_i$. Only for dimension $n = 2, 4$ or 8 the family of HR matrices consists of $N - 1$ matrices. For $n = 2$ there is one matrix:

$$A_1 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

For $n = 4$ there are three HR matrices with integer entries:

$$A_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, A_2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}, A_3 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix}.$$

For $N = 8$ we have seven HR matrices with elements $0, \pm 1$ [7]. So far HR matrices are applied in electronics [22]; in Space-Time Block Coding (STBC) and orthogonal design [23], also in signal processing [24] and Hamiltonian Neural Nets [25].

If one curve is described by a set of representation points $\{(x_i, y_i), i = 1, 2, \dots, n\}$ monotonic in coordinates x_i , then HR matrices combined with

identity matrix I_N are used to build an orthogonal and discrete Hurwitz - Radon Operator (OHR). For nodes $(x_1, y_1), (x_2, y_2)$ OHR M of dimension $N = 2$ is constructed [26]:

$$B = (x_1 \cdot I_2 + x_2 \cdot A_1)(y_1 \cdot I_2 - y_2 \cdot A_1) = \begin{bmatrix} x_1 & x_2 \\ -x_2 & x_1 \end{bmatrix} \begin{bmatrix} y_1 & -y_2 \\ y_2 & y_1 \end{bmatrix}, M = \frac{1}{x_1^2 + x_2^2} B,$$

$$M = \frac{1}{x_1^2 + x_2^2} \begin{bmatrix} x_1 y_1 + x_2 y_2 & x_2 y_1 - x_1 y_2 \\ x_1 y_2 - x_2 y_1 & x_1 y_1 + x_2 y_2 \end{bmatrix}. \quad (1)$$

Matrix M in (1) is found as a solution of equation:

$$\begin{bmatrix} a & b \\ -b & a \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}. \quad (2)$$

For nodes $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)$, monotonic in x_i , OHR of dimension $N = 4$ is constructed [26]:

$$M = \frac{1}{x_1^2 + x_2^2 + x_3^2 + x_4^2} \begin{bmatrix} u_0 & u_1 & u_2 & u_3 \\ -u_1 & u_0 & -u_3 & u_2 \\ -u_2 & u_3 & u_0 & -u_1 \\ -u_3 & -u_2 & u_1 & u_0 \end{bmatrix} \quad (3)$$

where $u_0 = x_1 y_1 + x_2 y_2 + x_3 y_3 + x_4 y_4, u_1 = -x_1 y_2 + x_2 y_1 + x_3 y_4 - x_4 y_3,$

$$u_2 = -x_1 y_3 - x_2 y_4 + x_3 y_1 + x_4 y_2, u_3 = -x_1 y_4 + x_2 y_3 - x_3 y_2 + x_4 y_1.$$

Matrix M in (3) is found as a solution of equation:

$$\begin{bmatrix} a & b & c & d \\ -b & a & -d & c \\ -c & d & a & -b \\ -d & -c & b & a \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}. \quad (4)$$

For nodes $(x_1, y_1), (x_2, y_2), \dots, (x_8, y_8)$, monotonic in x_i , OHR of dimension $N = 8$ is built [26] similarly as (1) or (3). Note that OHR operators M (1)-(3) satisfy the condition of interpolation

$$M \mathbf{x} = \mathbf{y} \quad (5)$$

for $\mathbf{x} = (x_1, x_2, \dots, x_N)^T \in \mathbf{R}^N, \mathbf{x} \neq 0, \mathbf{y} = (y_1, y_2, \dots, y_N)^T \in \mathbf{R}^N, N = 2, 4$ or 8.

If one curve is described by a set of nodes $\{(x_i, y_i), i = 1, 2, \dots, n\}$ monotonic in coordinates y_i , then HR matrices combined with identity matrix I_N are used to build an orthogonal and discrete reverse Hurwitz - Radon Operator (reverse OHR) M^{-1} . If matrix M in (1)-(3) has form:

$$M = \frac{1}{\sum_{i=1}^N x_i^2} (u_0 \cdot I_N + D),$$

where matrix D with elements u_1, \dots, u_{N-1} and zero diagonal, then reverse OHR M^{-1} is given by:

$$M^{-1} = \frac{1}{\sum_{i=1}^N y_i^2} (u_0 \cdot I_N - D). \quad (6)$$

Note that reverse OHR operator (6) satisfies the condition of interpolation

$$M^{-1} \mathbf{y} = \mathbf{x}. \quad (7)$$

for $\mathbf{x} = (x_1, x_2, \dots, x_N)^T \in \mathbf{R}^N$, $\mathbf{y} = (y_1, y_2, \dots, y_N)^T \in \mathbf{R}^N$, $\mathbf{y} \neq 0$, $N = 2, 4$ or 8 .

3.2 MHR Method (Basic Version)

Key question looks as follows: how compute coordinates of points settled between interpolation nodes? A set of nodes is the only information about curve in basic version of MHR method. On a segment of a line every number "c" situated between "a" and "b" is described by a linear (convex) combination $c = \alpha a + (1 - \alpha)b$ for

$$\alpha = \frac{b - c}{b - a} \in [0; 1]. \quad (8)$$

When the nodes are monotonic in coordinates x_i , average OHR operator M_2 of dimension $N = 2, 4$ or 8 is constructed as follows [7,26]:

$$M_2 = \alpha \cdot M_0 + (1 - \alpha) \cdot M_1 \quad (9)$$

with the operator M_0 built (1)-(3) by "odd" nodes $(x_1 = a, y_1), (x_3, y_3), \dots, (x_{2N-1}, y_{2N-1})$ and M_1 built (1)-(3) by "even" nodes $(x_2 = b, y_2), (x_4, y_4), \dots, (x_{2N}, y_{2N})$. Having the operator M_2 for coordinates $x_i < x_{i+1}$ it is possible to reconstruct the second coordinates of points (x, y) in terms of the vector C defined with

$$c_i = \alpha \cdot x_{2i-1} + (1 - \alpha) \cdot x_{2i}, i = 1, 2, \dots, N \quad (10)$$

as $C = [c_1, c_2, \dots, c_N]^T$. The required formula is similar to (5):

$$Y(C) = M_2 \cdot C \quad (11)$$

in which components of vector $Y(C)$ give the second coordinate of the points (x, y) corresponding to the first coordinate, given in terms of components of the vector C . On the other hand, having the operator M_2^{-1} for coordinates $y_i < y_{i+1}$ it is possible to reconstruct the first coordinates of points (x, y) [7, 26]:

$$M_2^{-1} = \alpha \cdot M_0^{-1} + (1 - \alpha) \cdot M_1^{-1}, c_i = \alpha \cdot y_{2i-1} + (1 - \alpha) \cdot y_{2i},$$

$$X(C) = M_2^{-1} \cdot C. \quad (12)$$

Contour of the object is constructed with several number of curves. Calculation of unknown coordinates for contour points using (8)-(12) is called by author the method of Hurwitz - Radon Matrices (MHR). Here is the application of basic MHR method for functions $f(x) = 2/x$ (five nodes equidistance in first coordinate: $x = 0.4, 0.7, 1.0, 1.3, 1.6$) and $f(x) = 1/(1 + 5x^2)$ with five nodes for $x = -1, -0.5, 0, 0.5, 1$.

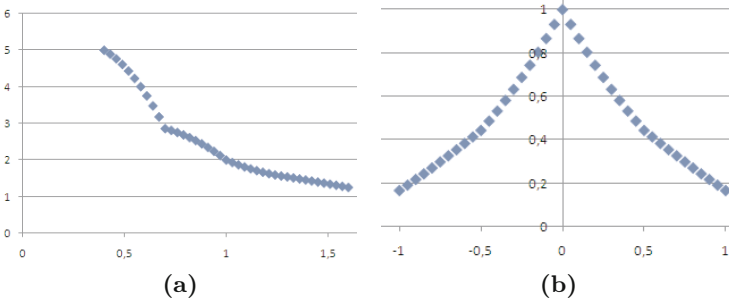


Fig. 3. Thirty six interpolated points of functions $f(x) = 2/x$ (a) and $f(x) = 1/(1 + 5x^2)$ (b) using basic MHR method with 5 nodes

Basic version of MHR method preserves monotonicity and symmetry (Fig. 3b) of the graphs.

3.3 MHR Method with Parameter k

The curve $y = 2/x$ reconstructed by basic version of MHR method (Fig. 3a) looks not quite accurate. For better reconstruction of the curve, appropriate $k \in (0; 2]$ in MHR method with parameter k is calculated:

$$M_2 = \alpha^k \cdot M_0 + (1 - \alpha^k) \cdot M_1 \quad (13)$$

or

$$M_2^{-1} = \alpha^k \cdot M_0^{-1} + (1 - \alpha^k) \cdot M_1^{-1} \quad (14)$$

For $k = 1$ MHR method (13-14) presents a basic version (9,12). In the case of $k > 2$ author's experiments confirm that models differ from the curves considerably. Choice of parameter k is connected with comparison of precise values w_i for function $f(x) = 2/x$ in control points p_i , situated in the middle between interpolation nodes ($\alpha = 0.5$):

$$p_i = \frac{1}{2}(x_i + x_{i+1}), w_i = f(p_i) = \frac{2}{p_i},$$

and values in control points p_i computed by MHR method. Control points are settled in the middle between interpolation nodes, because interpolation error of MHR method is the biggest [6]. Choice of rank k is done by criterion: difference between precise values w_i and values reconstructed by MHR method is the smallest. Control points p_i in this example are established for $p_i = 0.55, 0.85, 1.15, 1.45$. Four values of the curve are compared for various parameter $k \in (0; 2]$. The best result is calculated for $k = 1.565$:

$$|w_1 - 3.637| + |w_3 - 1.624| + |w_2 - 2.278| + |w_4 - 1.322| = 0.248,$$

whereas basic version ($k = 1$) gives worse result:

$$|w_1 - 4.23| + |w_3 - 1.709| + |w_2 - 2.532| + |w_4 - 1.398| = 0.822,$$

Reconstruction of the curve $y = 2/x$ by MHR method (13) with parameter $k = 1.565$ looks as follows:

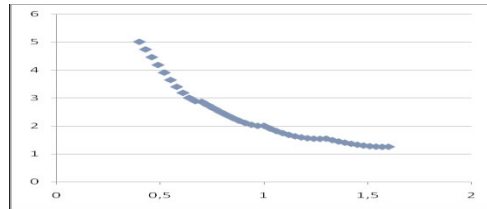


Fig. 4. The curve $y = 2/x$ modeled via MHR method for $k = 1.565$ and five nodes together with 36 reconstructed points

Fig. 4 represents the curve $y = 2/x$ more precisely than Fig. 3a. Convexity of reconstructed curve is very important factor in MHR method. Appropriate choice of parameter k is connected with regulation and controlling of convexity: model of the curve (Fig. 4) preserves monotonicity and convexity.

3.4 MHR Method for Equidistance Nodes

Assume that there is odd number of interpolation nodes $(x_1, y_1), (x_2, y_2), \dots, (x_{2k+1}, y_{2k+1})$ in MHR method ($k = 2, 3, 4, \dots, k = \text{const.}$) and all coordinates x_i or all coordinates y_i are equidistance (a fixed step of coordinates x_i or y_i). For example dealing with coordinate x_i we have the condition of proportion for first and second half of nodes ($n = 2$):

$$\forall i = 2, \dots, k : \frac{x_{k+1} - x_i}{x_{k+1} - x_1} = \frac{x_{2k+1} - x_{k+i}}{x_{2k+1} - x_{k+1}} = p_{i-1}. \quad (15)$$

Values $p_1 > \dots > p_{k-1} \in (0; 1)$ with $p_0 = 1$ and $p_k = 0$ are crucial in the process of interpolation. Let $M_i (i = 0, 1, 2, \dots, k)$ is OHR operator of

dimension $n = 2$ constructed (11) for nodes (x_{i+1}, y_{i+1}) and (x_{k+i+1}, y_{k+i+1}) . Average OHR operator M_{k+1} is built as follows:

$$M_{k+1} = \sum_{i=0}^k s_i \cdot M_i. \quad (16)$$

Average OHR operator M_2 in basic version (9) is calculated as (16) for $k = 1$ and $p_1 = 0$. Coefficients s_i are computed:

$$s_i = \frac{(\alpha - p_0)(\alpha - p_1) \dots (\alpha - p_{i-1})(\alpha - p_{i+1}) \dots (\alpha - p_k)}{(p_i - p_0)(p_i - p_1) \dots (p_i - p_{i-1})(p_i - p_{i+1}) \dots (p_i - p_k)}, \quad (17)$$

$$s_i = \frac{\prod_{j=0, j \neq i}^k (\alpha - p_j)}{\prod_{j=0, j \neq i}^k (p_i - p_j)}, \sum_{i=0}^k s_i = 1$$

for any coordinate c_1 situated between x_1 and x_{k+1} (first half of nodes) as follows:

$$c_1 = \alpha \cdot x_1 + \beta \cdot x_{k+1} \text{ for } 0 \leq \beta = 1 - \alpha \leq 1,$$

$$\alpha = \frac{x_{k+1} - c_1}{x_{k+1} - x_1} \in [0; 1]. \quad (18)$$

Vector of second coordinates $Y(C) = [y(c_1), y(c_2)]^T$ is calculated:

$$Y(C) = M_{k+1} \cdot \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = M_{k+1} \cdot (\alpha \begin{bmatrix} x_1 \\ x_{k+1} \end{bmatrix} + (1 - \alpha) \begin{bmatrix} x_{k+1} \\ x_{2k+1} \end{bmatrix}). \quad (19)$$

Here is the example of average operator (16) for five nodes equidistance in coordinate x_i : $(x_1, y_1), (x_2, y_2), \dots, (x_5, y_5)$, $k = 2, p_2 = 0, p_1 = 1/2, p_0 = 1$.

$$M_0 = \frac{1}{x_1^2 + x_3^2} \begin{bmatrix} x_1 y_1 + x_3 y_3 & x_3 y_1 - x_1 y_3 \\ x_1 y_3 - x_3 y_1 & x_1 y_1 + x_3 y_3 \end{bmatrix}$$

$$M_1 = \frac{1}{x_2^2 + x_4^2} \begin{bmatrix} x_2 y_2 + x_4 y_4 & x_4 y_2 - x_2 y_4 \\ x_2 y_4 - x_4 y_2 & x_2 y_2 + x_4 y_4 \end{bmatrix},$$

$$M_2 = \frac{1}{x_3^2 + x_5^2} \begin{bmatrix} x_3 y_3 + x_5 y_5 & x_5 y_3 - x_3 y_5 \\ x_3 y_5 - x_5 y_3 & x_3 y_3 + x_5 y_5 \end{bmatrix}.$$

$$s_0 = \frac{(\alpha - 0)(\alpha - 0.5)}{(1 - 0)(1 - 0.5)}, s_1 = \frac{(\alpha - 0)(\alpha - 1)}{(0.5 - 0)(0.5 - 1)}, s_2 = \frac{(\alpha - 1)(\alpha - 0.5)}{(0 - 1)(0 - 0.5)},$$

$$\sum_{i=0}^2 s_i = 2\alpha(\alpha - \frac{1}{2}) - 4\alpha(\alpha - 1) + 2(\alpha - 1)(\alpha - \frac{1}{2}) = 1,$$

$$M_3 = 2\alpha(\alpha - \frac{1}{2})M_0 - 4\alpha(\alpha - 1)M_1 + 2(\alpha - 1)(\alpha - \frac{1}{2})M_2.$$

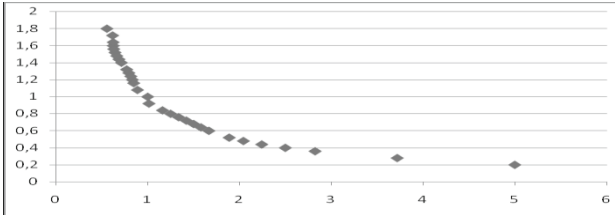


Fig. 5. Twenty two interpolated points of function $f(x) = 1/x$ using MHR method with 9 equidistance nodes

Here is the application of MHR method with equidistance nodes for function $f(x) = 1/x$ and nine nodes equidistance in second coordinate: $y = 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8$.

MHR method for equidistance nodes requires the coefficients s_i (17) that are computed in similar way like Lagrange interpolation polynomial.

4 Conclusion

The method of Hurwitz-Radon Matrices leads to contour interpolation and shape reconstruction depending on the number and location of representation points. No characteristic features of curve are important in MHR method: failing to be differentiable at any point, the Runge's phenomenon or differences from the shape of polynomials. These features are very significant for classical polynomial interpolations. MHR method gives the possibility of reconstruction a curve consists of several parts, for example closed curve (contour). The only condition is to have a set of nodes for each part of a curve or contour according to assumptions in MHR method. Shape representation and curve reconstruction by MHR method is connected with possibility of changing the nodes coordinates and reconstruction of new curve or contour for new set of nodes, no matter what shape of curve or contour is to be reconstructed. Main features of MHR method are: accuracy of shape reconstruction depending on number of nodes and method of choosing nodes; reconstruction of curve consists of L points is connected with the computational cost of rank $O(L)$ (18); MHR method preserves monotonicity and symmetry of the graphs, but convexity not always (selection of parameter k).

Future works are connected with: geometrical transformations of contour (translations, rotations, scaling)- only nodes are transformed and new curve (for example contour of the object) for new nodes is reconstructed, possibility to apply MHR method to three-dimensional curves (26) and connection MHR method with object recognition.

References

1. Ballard, D.H.: Computer Vision. Prentice Hall, New York (1982)
2. Tadeusiewicz, R., Flasiński, M.: Image Recognition. PWN, Warsaw (1991)
3. Saber, E., Yaowu, X., Murat Tekalp, A.: Partial shape recognition by submatrix matching for partial matching guided image labeling. *Pattern Recognition* 38, 1560–1573 (2005)
4. Sebastian, T.B., Klein, P.N.: On aligning curves. *IEEE Trans. Pattern Anal. Mach. Intell.* 25(1), 116–124 (2003)
5. Liu, T., Geiger, D.: Aroximate tree matching and shape similarity. In: *Int. Conf. Computer Vision*, Corfu, Greece (1999)
6. Choraś, R.S.: *Komputerowa wizja*. Exit, Warszawa (2005)
7. Jakóbczak, D., Kosiński, W.: Application of Hurwitz - Radon Matrices in Monochromatic Medical Images Decompression. In: Kowalczyk, Z., Wiszniewski, B. (eds.) *Intelligent Data Mining in Diagnostic Purposes: Automatics and Informatics*, PWN, Gdansk, pp. 389–398 (2007)
8. Latecki, L.J., Lakaemper, R.: Convexity Rule for Shape Decomposition Based on Discrete Contour Evolution. *Computer Vision and Image Understanding* 3(73), 441–454 (1999)
9. Cierniak, R.: *Tomografia komputerowa*. Exit, Warszawa (2005)
10. Soussen, C., Mohammad-Djafari, A.: Polygonal and Polyhedral Contour Reconstruction in Computed Tomography. *IEEE Transactions on Image Processing* 11(13), 1507–1523 (2004)
11. Tang, K.: Geometric Optimization Algorithms in Manufacturing. *Computer – Aided Design & Applications* 2(6), 747–757 (2005)
12. Kozera, R.: *Curve Modeling via Interpolation Based on Multidimensional Reduced Data*. Silesian University of Technology Press, Gliwice (2004)
13. Schumaker, L.L.: *Spline functions: basic theory*. Cambridge Mathematical Library (2007)
14. Dejdumrong, N.: A shape preserving verification techniques for parametric curves. In: *Computer Graphics, Imaging and Visualization, CGIV 2007*, pp. 163–168 (2007)
15. Dyn, N., Levin, D., Gregory, J.A.: A 4-point interpolatory subdivision scheme for curve design. *Comput. Aided. Geom. Design* 4, 257–268 (1987)
16. Rogers, D.F.: *An Introduction to NURBS with Historical Perspective*. Morgan Kaufmann Publishers, San Francisco (2001)
17. Kiciak, P.: *Podstawy modelowania krzywych i powierzchni. Zastosowania w grafice komputerowej*. WNT, Warszawa (2005)
18. Jakóbczak, D.: Curve Interpolation Using Hurwitz-Radon Matrices. *Polish Journal of Environmental Studies* 3B(18), 126–130 (2009)
19. Dahlquist, G., Björck, A.: *Numerical Methods*. Prentice Hall, New York (1974)
20. Ralston, A.: *A First Course in Numerical Analysis*. McGraw-Hill Book Company, New York (1965)
21. Eckmann, B.: Topology, Algebra, Analysis- Relations and Missing Links. *Notices of the American Mathematical Society* 5(46), 520–527 (1999)
22. Citko, W., Jakóbczak, D., Sieńko, W.: On Hurwitz - Radon Matrices Based Signal Processing. In: *Workshop Signal Processing at Poznan University of Technology* (2005)

23. Tarokh, V., Jafarkhani, H., Calderbank, R.: Space-Time Block Codes from Orthogonal Designs. *IEEE Transactions on Information Theory* 5(45), 1456–1467 (1999)
24. Sieńko, W., Citko, W., Wilamowski, B.: Hamiltonian Neural Nets as a Universal Signal Processor. In: 28th Annual Conference of the IEEE Industrial Electronics Society IECON (2002)
25. Sieńko, W., Citko, W.: Hamiltonian Neural Net Based Signal Processing. In: The International Conference on Signal and Electronic System ICSES (2002)
26. Jakóbczak, D.: 2D and 3D Image Modeling Using Hurwitz-Radon Matrices. *Polish Journal of Environmental Studies* 4A(16), 104–107 (2007)

Video Quality Assessment Using the Combined Full-Reference Approach

Krzysztof Okarma

West Pomeranian University of Technology
Faculty of Electrical Engineering
Department of Signal Processing and Multimedia Engineering
26. Kwietnia 10, 71-126 Szczecin, Poland
e-mail: okarma@zut.edu.pl

Summary. In this paper the new combined video quality metric is proposed, which may be useful for the quality assessment of the compressed video files, especially transmitted using wireless channels. The proposed metric is the weighted combination of three state-of-the-art image quality metrics, which are well correlated with the subjective evaluations. A simple extension of those metrics for the video quality assessment is the averaging of their values for all video frames. Nevertheless, such approach may not lead to satisfactory results for all types of distortions. In this paper the typical distortions introduced during the wireless video transmission have been analyzed using the 160 files available as the LIVE Wireless Video Quality Assessment Database together with the results of subjective quality evaluation. Obtained results are promising and the proposed metric is superior to each of the analyzed ones in the aspect of the linear correlation with subjective scores.

1 Introduction

Automatic image and video quality assessment is still one of the major fields of interest of the researchers related to the multimedia systems. Starting from the Mean Squared Error (MSE) traditionally used as the measure of the similarity for the grayscale images, a lot of research has been done in order to propose some image quality assessment methods, which are much better correlated with the way human perceive images and the various types of distortions, which may affect them. The necessity of the reliable quality image and video assessment is still the same: the development of some new processing methods, e.g. filtration of lossy compression, requires the methods of the quality estimation in order to measure the visual results of processing.

2 Methods of Image Quality Assessment

The basic ideas of the video quality assessment are naturally based on the methods used for the quality assessment of the still images. The simplest way of their extension is the frame-by-frame calculation of the image quality index and further averaging of the obtained results. Despite the fact that such approach does not concern in general to some of the properties of the Human Visual System (HVS) e.g. related to temporal masking, eye movements etc., it can be successfully applied due to its simplicity. The main requirement is the good correlation of the quality metric with subjective evaluations.

Another complication is the fact that the most of the image quality metrics, even the most recent ones, are not designed for the assessment of the color images. Since, their extension into the multichannel versions is not straightforward [1, 2], their applicability for the color video quality assessment is still an open field of research.

The image and video quality assessment can be performed using the objective metrics or subjective evaluations. The first group is much more desired, since the scalar value representing the image or video quality can easily be used for the optimization of the image and video processing algorithms (e.g. lossy compression). Another advantage of the objective metrics is the possibility of their fast computation in contrast to the subjective scores, which require the time-consuming human interaction. Nevertheless, subjective evaluation, mainly Mean Opinion Scores (MOS), can be used for the verification of some developed objective metrics.

All objective image and video quality assessment methods can be divided into three major groups: full-reference, reduced-reference [3] and no-reference (blind) ones [4]. The most metrics, characterized by high universality, belong to the first group, where the full knowledge of the ideal reference image without any distortions is required. Blind image quality assessment methods are rather specialized and sensitive on only one or two types of distortions e.g. blurring [5, 6] or blocking artifacts caused by the JPEG compression [7, 8].

The traditional approach to the objective image quality assessment is based on the Mean Squared Error and some similar metrics such as e.g. Peak Signal to Noise Ratio (PSNR) [9, 10]. Nevertheless, their main drawback is poor correlation with subjective perception of distortions. A great impulse for the development of some new methods of the image quality assessment has been the definition of the Universal Image Quality Index [11], further extended into Structural Similarity (SSIM) [12]. This metric became probably the most popular one due to its low computational complexity and good correlation with human perception of typical image distortions. The overall SSIM value is computed as the average value of the local indexes obtained using the sliding window (11 × 11 pixels Gaussian window is default) according to the formula:

$$SSIM = \frac{(2\bar{x}\bar{y} + C_1) \cdot (2\sigma_{xy} + C_2)}{(\sigma_x^2 + \sigma_y^2 + C_1) \cdot (\bar{x}^2 + \bar{y}^2 + C_2)}, \quad (1)$$

where C_1 and C_2 are small values preventing the possible division by zero for dark large regions of the same color. The default values suggested by the authors of the paper [12] are $C_1 = (0.01 \times 255)^2$ and $C_2 = (0.03 \times 255)^2$. The symbols \bar{x} and \bar{y} stand for the average values of the original and distorted image respectively, σ_x^2 , σ_y^2 and σ_{xy} are the respective variances and the covariance for the fragments of both images inside the current window. The default window can be changed, especially its size, depending on the resolution of images [13].

Further extension of this metric has been proposed as the multi-scale version [14] operating over a dyadic pyramid. The luminance (l), contrast (c) and structure (s) factors computed for each scale are weighted using the specified exponents values. The final MS-SSIM value can be calculated as

$$\text{MS-SSIM}(x, y) = [l_M(x, y)]^{\alpha_M} \cdot \prod_{j=1}^M [c_j(x, y)]^{\beta_j} [s_j(x, y)]^{\gamma_j} , \quad (2)$$

where M denotes the highest scale obtained after $M - 1$ iterations of low-pass filtering and downsampling such obtained image by a factor of 2.

One of the latest trends in the image quality assessment is the use of the Singular Value Decomposition (SVD). The first approach has been the M_{SVD} metric [15, 16] and in some of the latest publications the reflection factors [17] and R-SVD metric [18] have been proposed. The last two metrics are based on similar assumptions but the R-SVD leads to better results. It can be calculated as

$$\text{R-SVD} = \frac{\sqrt{\sum_{i=1}^m (d_i - 1)^2}}{\sqrt{\sum_{i=1}^m (d_i + 1)^2}} , \quad (3)$$

where d_i denotes the singular values of the reference matrix $\hat{R} = \hat{U}A V^T$. The matrices U , S and V^T are the result of the SVD decomposition of the original image A , A stands for the identity matrix and \hat{U} , \hat{S} and \hat{V}^T of the distorted one respectively. Instead of the right singular matrix of the distorted image \hat{V}^T , the original matrix V^T can be used and the left singular matrix \hat{U} can be computed as

$$\hat{U}_i = \begin{cases} 0 & \text{if } \hat{s}_i = 0 \\ \hat{A} \cdot V_i / \hat{s}_i & \text{otherwise} \end{cases} , \quad (4)$$

where \hat{s}_i are the singular values of the matrix representing the distorted image (\hat{A}).

Another modern image quality metric, well correlated with human perception is the Visual Information Fidelity (VIF) metric [19, 20]. This metric is

based on the wavelet decomposition but the lower complexity pixel domain version can also be used. The general definition of this metric is

$$VIF = \frac{\sum_{j=0}^S \sum_{i=1}^{M_j} I(c_{i,j}; f_{i,j})}{\sum_{j=0}^S \sum_{i=1}^{M_j} I(c_{i,j}; e_{i,j})}, \quad (5)$$

where S denotes the number of sub-bands (scales for the pixel domain version), M_j stands for the number of blocks at j -th sub-band (scale) and $I(x; y)$ denotes the mutual information between x and y . The denominator and numerator are interpreted as the information that vision extracts from the reference and distorted images, assuming that c is a block vector at a specified location in the reference image, e is the perception of block c by a human viewer with additive noise n , and f is the perception of distorted block c [20].

3 Combined Metric and Its Verification

One of the most desired properties of an ideal image and video quality metric is its linear correlation with subjective evaluations by human observers. It means that it can be calculated with inside nonlinear functions but the final results should be well correlated with the human perception without additional mapping. Some of the results presented by various researchers, with the correlation values even as high as 0.99 are obtained using such nonlinear mapping, usually using the logistic function, suggested by the Video Quality Experts Group (VQEG) [21]. The drawback is that the obtained coefficients of the logistic function are usually different for each type of distortions present in a given database, so their proper choice requires additional optimization [22]. In such sense the universality of such approach seems to be doubtful.

The modern image quality assessment methods discussed above have different properties and their good correlation with the Differential Mean Opinion Scores can be obtained only by the use of the logistic function for the nonlinear mapping. Nevertheless, the combination of three of them (MS-SSIM, VIF and R-SVD) can be used as the weighted product in order to obtain the higher values of the Pearson linear correlation coefficient. The verification of such approach for the still images from two commonly used image quality assessment databases has been presented in the earlier paper [23] leading to the increase of the linear correlation, calculated for 17 types of distortions, from about 0.78 for the MS-SSIM to 0.86 for the combined metric.

The proposed extension of the metrics for the video quality assessment is based on the weighting of the metrics discussed in Section 2. The general

form of the combined video quality metric, applied using the frame-by-frame approach (i denote the frame number), can be expressed as

$$CVQM = \left(\sum_i SSIM \right)^a \cdot \left(\sum_i MS-SSIM \right)^b \cdot \left(\sum_i VIF \right)^c \cdot \left(\sum_i R-SVD \right)^d . \tag{6}$$

The verification of its properties and the choice of the proper values of the exponents has been performed using the LIVE Wireless Video Quality Assessment Database [24, 25] containing 160 files contaminated by wireless channel-specific distortions for H.264 AVC compressed videos with 10 reference videos and the subjective Mean Opinion Scores delivered by over 30 viewers.

The Pearson linear correlation coefficients between MOS and the proposed combined metric have been computed for the combinations of the four discussed metrics. It is worth noticing that the obtained values of the exponents should not be treated as the exact ones, since the video quality assessment is in fact the quality estimation and the ideal quality metric does not exist. The first conclusion, different as for the still images, is related to the fact that the application of the R-SVD metric does not improve the obtained results if the other three metrics are used, so it can be removed from the metric ($d = 0$).

Further computations have been performed for the weighted combinations of two metrics in order to verify the necessity of each of the first three factors of the formula (6). The results are presented in Table 1 and the comparison of the linear correlation coefficients for each metric and the proposed one is presented in Table 2. The obtained nearly optimal values of the exponents are: $a = -2.9$, $b = 9.6$, $c = 0.7$ (and $d = 0$ as mentioned above). The scatter plots of the results obtained using the mentioned 160 files for each of the metrics applied separately and for the proposed combined metric are shown in Figs. 1 and 2.

Table 1. Pearson linear correlation coefficients obtained using the weighted combinations of two metrics

SSIM and MS-SSIM	SSIM and VIF	VIF and MS-SSIM
0.9631	0.9477	0.9677

Table 2. Pearson linear correlation coefficients obtained during experiments for the LIVE Wireless Video Quality Assessment Database

SSIM	MS-SSIM	VIF	R-SVD	Proposed
0.8578	0.8532	0.9447	0.8287	0.9734

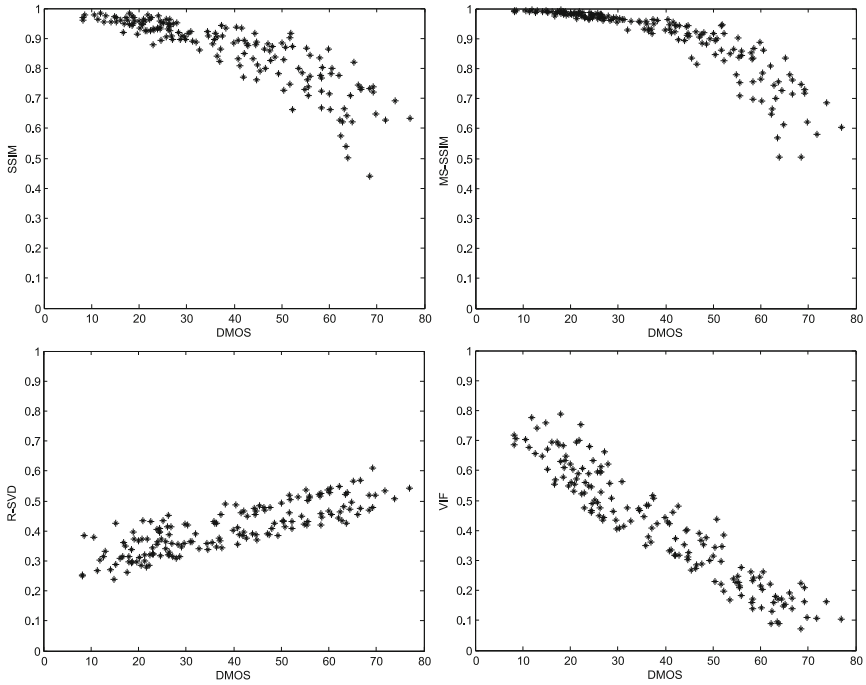


Fig. 1. Scatter plots of the mean values of four analyzed metrics versus the DMOS values for the 160 files from the LIVE Wireless Video Quality Assessment Database

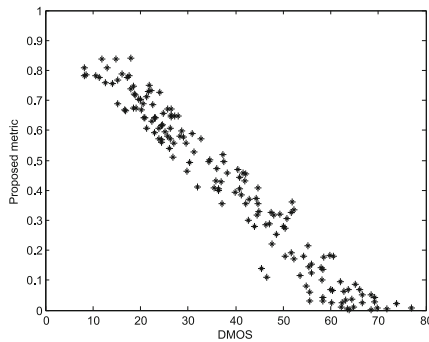


Fig. 2. Scatter plots of the mean values of the proposed combined metric versus the DMOS values for the 160 files from the LIVE Wireless Video Quality Assessment Database

4 Conclusions

The main advantage of the proposed video quality assessment method based on the nonlinear combination of some modern metrics is high linear

correlation with the subjective assessment. Although the database used in the experiments contains only the video files with wireless distortions, the obtained results are promising and can be further verified using some other types of contaminations. In that sense, the inclusion of the metric based on the Singular Value Decomposition may be useful. Nevertheless, planned further experiments require additional calculations using another video database containing the video files with other types of contaminations.

Possible directions of further research include also the application of some statistical methods [26] in order to increase the processing speed as well as the extension towards the color video quality assessment.

References

1. Toet, A., Lucassen, M.P.: A new universal colour image fidelity metric. *Displays* 24(4-5), 197–207 (2003)
2. Okarma, K.: Colour image quality assessment using Structural Similarity index and Singular Value Decomposition. In: Bolc, L., Kulikowski, J.L., Wojciechowski, K. (eds.) *ICCVG 2008*. LNCS, vol. 5337, pp. 55–65. Springer, Heidelberg (2009)
3. Wang, Z., Simoncelli, E.: Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In: *Proc. Human Vision and Electronic Imaging Conf. Proceedings of SPIE*, vol. 5666, pp. 149–159 (2005)
4. Li, X.: Blind image quality assessment. In: *Proc. IEEE Int. Conf. Image Proc.*, pp. 449–452 (2002)
5. Marziliano, P., Dufaux, F., Winkler, S., Ebrahimi, T.: A no-reference perceptual blur metric. In: *Proc. IEEE Int. Conf. Image Proc.*, pp. 57–60 (2002)
6. Ong, E.P.: Lin, Lu.W., Yang, Z., Yao, S., Pan, F., Jiang, L., Moschetti, F.: A no-reference quality metric for measuring image blur. In: *Proc. 7th Int. Symp. Signal Proc. and Its Applications*, pp. 469–472 (2003)
7. Wang, Z., Sheikh, H., Bovik, A.: No-reference perceptual quality assessment of JPEG compressed images. In: *Proc. IEEE Int. Conf. Image Proc.*, pp. 477–480 (2002)
8. Wang, Z., Bovik, A., Evans, B.: Blind measurement of blocking artifacts in images. In: *Proc. IEEE Int. Conf. Image Proc.*, pp. 981–984 (2000)
9. Eskicioglu, A., Fisher, P., Chen, S.: Image quality measures and their performance. *IEEE Trans. Comm.* 43(12), 2959–2965 (1995)
10. Eskicioglu, A.: Quality measurement for monochrome compressed images in the past 25 years. In: *Proc. Int. Conf. Acoust. Speech Signal Proc.*, pp. 1907–1910 (2000)
11. Wang, Z., Bovik, A.: A universal image quality index. *IEEE Signal Proc. Letters* 9(3), 81–84 (2002)
12. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: From error measurement to Structural Similarity. *IEEE Trans. Image Proc.* 13(4), 600–612 (2004)
13. Okarma, K.: Two-dimensional windowing in the Structural Similarity index for the colour image quality assessment. In: Jiang, X., Petkov, N. (eds.) *CAIP 2009*. LNCS, vol. 5702, pp. 501–508. Springer, Heidelberg (2009)

14. Wang, Z., Simoncelli, E., Bovik, A.: Multi-Scale Structural Similarity for image quality assessment. In: Proc. 37th IEEE Asilomar Conf. on Signals, Systems and Computers (2003)
15. Shnayderman, A., Gusev, A., Eskicioglu, A.: A multidimensional image quality measure using Singular Value Decomposition. In: Proc. SPIE Image Quality and Syst. Perf., vol. 5294(1), pp. 82–92 (2003)
16. Shnayderman, A., Gusev, A., Eskicioglu, A.: An SVD-based gray-scale image quality measure for local and global assessment. *IEEE Trans. Image Proc.* 15(2), 422–429 (2006)
17. Mahmoudi-Aznavah, A., Mansouri, A., Torkamani-Azar, F., Eslami, M.: Image quality measurement besides distortion type classifying. *Optical Review* 16(1), 30–34 (2009)
18. Mansouri, A., Mahmoudi-Aznavah, A., Torkamani-Azar, F., Jahanshahi, J.A.: Image quality assessment using the Singular Value Decomposition theorem. *Optical Review* 16(2), 49–53 (2009)
19. Sheikh, H.R., Bovik, A., de Veciana, G.: An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans. Image Proc.* 14(12), 2117–2128 (2005)
20. Sheikh, H.R., Bovik, A.: Image information and visual quality. *IEEE Trans. Image Proc.* 15(2), 430–444 (2006)
21. VQEG Final report on the validation of objective models of video quality assessment (2003), <http://www.vqeg.org>
22. Sendashonga, M., Labeau, F.: Low complexity image quality assessment using frequency domain transforms. In: Proc. IEEE Int. Conf. Image Proc., pp. 385–388 (2006)
23. Okarma, K.: Combined full-reference image quality metric linearly correlated with subjective assessment. In: Rutkowski, L., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2010. LNCS (LNAI), vol. 6113, pp. 539–546. Springer, Heidelberg (2010)
24. Moorthy, A.K., Seshadrinathan, K., Soundararajan, R., Bovik, A.: Wireless video quality assessment: A study of subjective scores and objective algorithms. *IEEE Trans. Circuits and Systems for Video Technology* 20(4), 513–516 (2010)
25. Moorthy, A.K., Seshadrinathan, K., Soundararajan, R., Bovik, A.: LIVE Wireless Video Quality Assessment Database (2009), http://live.ece.utexas.edu/research/quality/live_wireless_video.html
26. Okarma, K., Lech, P.: A statistical reduced-reference approach to digital image quality assessment. In: Bolc, L., Kulikowski, J.L., Wojciechowski, K. (eds.) ICCVG 2008. LNCS, vol. 5337, pp. 43–54. Springer, Heidelberg (2009)

An Improved Self-embedding Algorithm: Digital Content Protection against Compression Attacks in Digital Watermarking

Pratheepan Yogarajah^{1,*}, Joan Condell¹, Kevin Curran¹,
Abbas Cheddad², and Paul McKeiv¹

¹ School of Computing and Intelligent Systems,
University of Ulster, UK

² Umea Centre for Molecular Medicine (UCMM),
Umea Universitet, Sweden

Summary. Lossy compression attacks in digital watermarking are one of the major issues in digital watermarking. Cheddad et al. proposed a robust secured self-embedding method which is resistant to a certain amount of JPEG compression. Our experimental results show that the self-embedding method is resistant to JPEG compression attacks and not resistant to other lossy compression attacks such as Block Truncation Coding (BTC) and Singular Value Decomposition (SVD). Therefore we improved Cheddad et al's. method to give better protection against BTC and SVD compression attacks.

1 Introduction

Protecting digital image content is an important task in image security. To protect the content, digital image watermarking techniques are applied. In watermarking the secret information called the watermark, is invisibly embedded into the host digital image. A general watermarking framework for content protection is presented in [1].

Watermarking techniques can be categorized into two types (spatial and frequency domain) according to the embedding process. Watermarking in the frequency domain is more robust than watermarking in the spatial domain [2], because the watermark information can be spread out over the entire image [3]. Commonly used frequency domain transforms include the Discrete Wavelet Transform (DWT), the Discrete Cosine Transform (DCT) and the Discrete Fourier Transform (DFT). However, DWT [4] has been used in digital image watermarking more frequently due to its excellent spatial localisation and multi-resolution characteristics, which are similar to the theoretical models of Human Visual System (HSV) [5].

* This work was part funded by the Invest NI Proof of Concept (PoC) fund.

The content of watermarked digital images can be easily attacked by using image processing operations such as lossy compression. Invisible watermarking requires a reasonable robustness against compression attacks. Lossy compression algorithms tend to remove invisible information that can be related to the watermark. Watermark robustness under image compression is an essential issue for image content protection. Therefore, watermarks should combine invisibility and robustness simultaneously.

Recently Cheddad et al [6] proposed a method to protect the digital image itself using a secured robust self-embedding technique. In their method, a halftoned version (black and white image) of the original image is used as watermark. The calculated watermark is embedded in the 2D Haar DWT of the original image and the watermarked image is obtained. Then the Wavelet-based Inverse Halftoning via De-convolution (WInHD) is used on the extracted watermark from the watermarked image to recover the approximation of the original image. This is a blind watermarking scheme as the original image is not needed for the recovery process, see Figure 1.

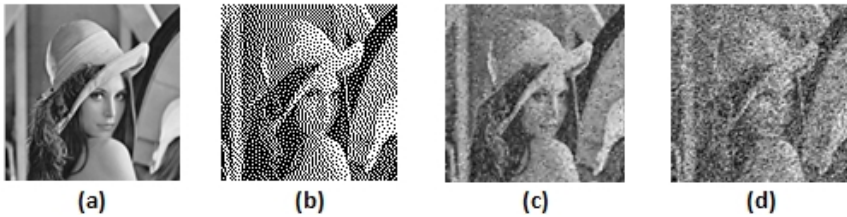


Fig. 1. (a) Original image, (b) halftoned binary image, (c) and (d) are recovered images from JPEG 95% and 85% quality compression attacks respectively

JPEG 2000 is one of the modern lossy compression methods and it is based on DWT. As the Cheddad et al. method is DWT based, it is resistant to JPEG compression attacks to a certain extent. They reported that their method is resilient to JPEG compression up to 80-75% [6]. There is no experimental results shown for other lossy compression techniques, such as Block Truncation Coding (BLC) and Singular Value Decomposition (SVD) etc. Our experimental results show that Cheddad et al's method is not robust to BTC and SVD lossy compression techniques. Therefore we improved Cheddad et al's method and experimental results prove that our method provides better recovery results on BTC and SVD compression attacks.

Our method is discussed in Section 2. The secure image encryption algorithm is explained in section 3. Sections 4 and 5 explain the results and provide discussion and conclusions respectively.

2 Our Method

In DWT, an image is decomposed into a set of basis functions namely low frequency band (LL), high-low frequency band (HL), low-high frequency band (LH) and high frequency band (HH), see Figure 2(a). The low frequency band is a lowpass approximation of the original image and includes most energy of the image. The other bands include edge components of horizontal, vertical and diagonal directions at different scales and resolutions respectively.

According to the energy distribution, LL, is the most important. Hence in DWT domain, watermarks should be embedded in the low frequency band [7]. Cheddad et al. also selected the low frequency band of the 1st-level 2D Haar DWT as their embedding area (i.e. LL_1) [6].

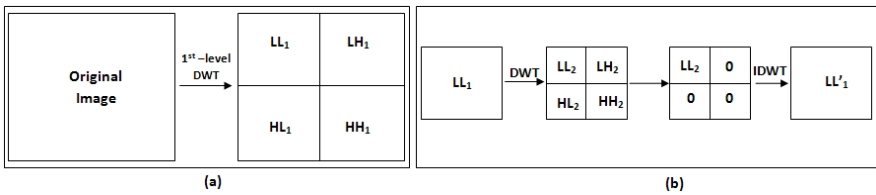


Fig. 2. (a) 1st-level DWT decomposition and (b) LL'_1 calculation

Common image processing procedures, which watermarked images may encounter, such as data compression, low-pass filtering and subsampling, tend to change the low frequency coefficients less than high frequency coefficients [7]. Therefore it is necessary to reduce the high frequency coefficient effects to make the watermarked image resistant to compression attacks.

DWT has applications in image processing, where typically the approach is to DWT an image, alter the transform coefficients (by thresholding or zeroing), and inverse DWT to regain an altered image that has been de-noised, or its edges sharpened or blurred. The zeroing high frequency coefficient technique is applied to the digital image watermarking application in [8].

We also applied the zeroing technique to improve the performance of Cheddad et al.'s method against lossy compression attacks. In our method the original LL_1 is further wavelet transformed and then three high frequency bands (LH_2 , HL_2 and HH_2 , excluding LL_2) are initialised to zeros and inverse wavelet transformed. We then obtain another LL'_1 from the process, see Figure 2(b). In our method, this newly calculated LL'_1 is used for embedding instead of LL_1 .

In our digital image content protection method, first the digital image, that needs to be protected against compressions attacks, is selected (say 256x256 'Lena' grayscale image). Then it is transformed to the DWT domain and LL'_1 is generated. The size of LL'_1 is half that of the original image. Therefore the original image is resized to be equal to the size of LL'_1 and Floyd's [9] error

diffusion digital halftoning technique is applied to obtain the watermark (i.e. a halftoned black and white image).

This watermark is embedded in LL'_1 using an encryption algorithm explained in section 3 and the inverse DWT is applied to generate the watermarked image. The embedded watermark is extracted from the compression attacked watermarked image using a decryption algorithm (again see Section 3). Finally the Wavelet-based Inverse Halftoning via De-convolution (WInHD) [10] is applied to the extracted watermark to recover the approximation of the original image.

Embedding the watermark into LL'_1 , instead of LL_1 , may decrease the watermarked image quality, see Figure 3(b), but the extracted watermark reliability is increased, see Figure 3(d). The BTC and SVD compression attacks are applied to watermarked images. Brief information on BTC and SVD given below.

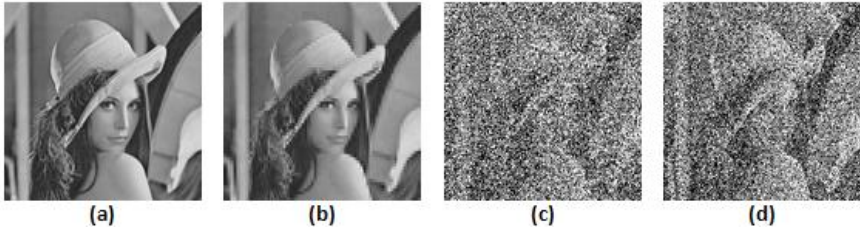


Fig. 3. (a) watermarked image using Cheddad et al's method, (b) watermarked image using our method, (c) recovered image from 4x4 blocks BTC compression attack using Cheddad et al's method and (d) recovered image from 4x4 blocks BTC compression attack using our method

Block Truncation Coding (BTC) is a lossy image compression technique. It divides the original images into blocks and then uses a quantizer to reduce the number of grey levels in each block while maintaining the same mean and standard deviation [11]. Sub blocks of 4x4 pixels allow compression of about 25%. Larger blocks allow greater compression however quality also reduces with the increase in block size due to the nature of the algorithm.

Singular Value Decomposition (SVD) is one of the most useful tools of linear algebra. It is a factorization and approximation technique which effectively reduces any matrix into a smaller invertible and square matrix. Using (SVD) for image compression can be a very useful tool to save storage space [12, 13].

3 Image Encryption Algorithm

This algorithm is explained based on [6] and the encryption algorithm is fully described in [14]. A hash function is more formally defined as the mapping of

bit strings of an arbitrary finite length to strings of fixed length [15]. Here we attempt to extend SHA-1 (the terminology and functions used as building blocks to form SHA-1 are described in the US Secure Hash Algorithm 1, [16]) to encrypt digital 2D data. The introduction of Fast Fourier Transform (FFT) forms together with the output of SHA-1 a strong image encryption setting. Let the key bit stream be $\lambda_{k,l}$ where the subscripts k and l denote the width and height after resizing the key's bit stream respectively, i.e., $8, M * N$, where M, N are the plain image's dimensions.

The FFT will operate on the DCT transform of $\lambda_{k,l}$ subject to Eq. 2.

$$f(u, v) = \frac{1}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} F(x, y) e^{-2\pi i(xu+yv)/N} \quad (1)$$

where $F(x, y) = DCT(\lambda_{k,l})$ satisfying Eq. (2). Note that for the transformation at the FFT and DCT levels we do not utilise all of the coefficients. Rather, we impose the following rule, which generates at the end a binary random-like map. Given the output of Eq. 1 we can derive the binary map straightforwardly as:

$$Map(x, y) = \begin{cases} 1 & \text{iff } f(u, v) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

This map takes the positive coefficients of the imaginary part to form the ON pixels in the map. Since the coefficients are omitted the reconstruction of the password phrase is impossible, hence the name Irreversible Fast Fourier Transform (IrFFT). In other words, it is a one way function which accepts initially a user password. This map finally is XORed with the binary version of each colour component separately. The core idea here is to transform these changes into the spatial domain where we can apply 2D-DCT and 2D-FFT that introduce the aforementioned sensitivity to the two dimensional space. As such, images can be easily encoded securely with password protection.

The watermark images are securely embedded using the encryption algorithm explained above. The decryption technique is also similar to the encryption algorithm and can be referred in [14].

4 Results and Discussion

In this section, we illustrate and evaluate the performance of the proposed method against JPEG, BTC and SVD compression attacks on grayscale images. Here we present experimental results using the image 'Lena' (256x256 pixels, grayscale). The 'Lena' image is shown in Figure 1(a).

Then the watermarked images are generated using Cheddad et al. and our method, see Figure 3(a) and (b). These watermarked images are attacked by JPEG, BTC and SVD compression techniques. The binary watermarks are extracted from attacked watermarked images and the approximation of the original image is recovered.

For quantitative evaluation, the PSNR (Peak Signal-to-Noise Ratio) is introduced to evaluate the performance between the original image and recovered image. The PSNR is defined as follows:

$$\text{PSNR} = 10 \log_{10} \left(\frac{255^2}{\text{MSE}} \right) \text{dB} \quad (3)$$

$$\text{MSE} = \sum_{i=1}^n \sum_{j=1}^m \frac{(a_{i,j} - b_{i,j})^2}{n * m} \quad (4)$$

where $m * n$ is the image size, $a_{i,j}$ and $b_{i,j}$ are the corresponding pixel values of cover and recovered images.

4.1 JPEG Compression Attack

From Figure 4 we see that both methods perform similarly against JPEG compression quality factors 75%, 85%, 90% and 95%.

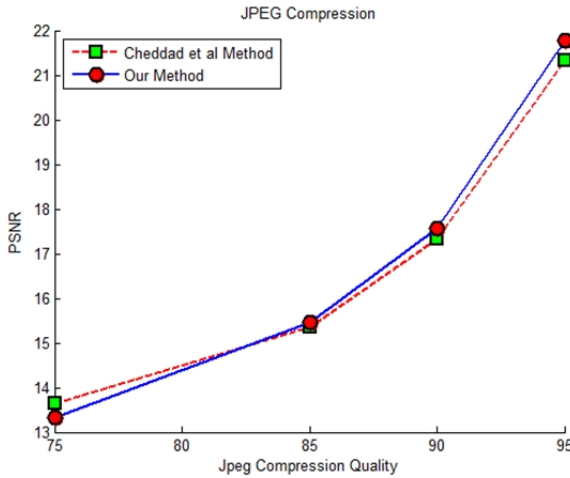


Fig. 4. Results of JPEG compression attack on 'Lena' image

4.2 BTC Compression Attack

Here 2x2, 4x4, 8x8 and 16x16 blocks are selected for experiments. When we apply 8x8 and 16x16 blocks BTC attacks on the watermarked image, the watermarked image becomes severely corrupted intensity of pixels. Therefore we could not see much difference in performance with 8x8 and 16x16 blocks BTC compression. Our method performed better when 2x2 and 4x4 blocks BTC compressions were applied to the watermarked images, see Figure 5.

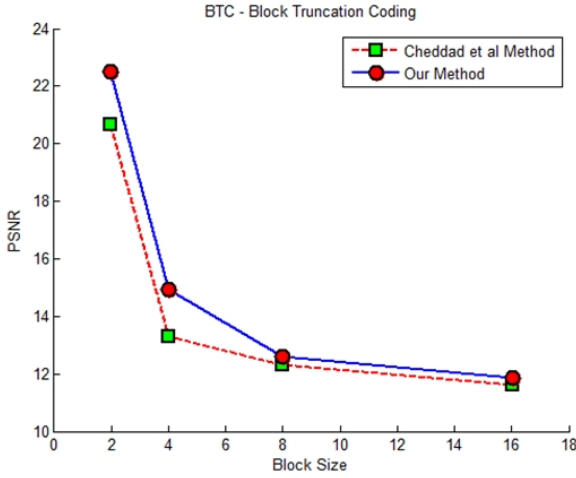


Fig. 5. Results of BTC compression attack on ‘Lena’ image

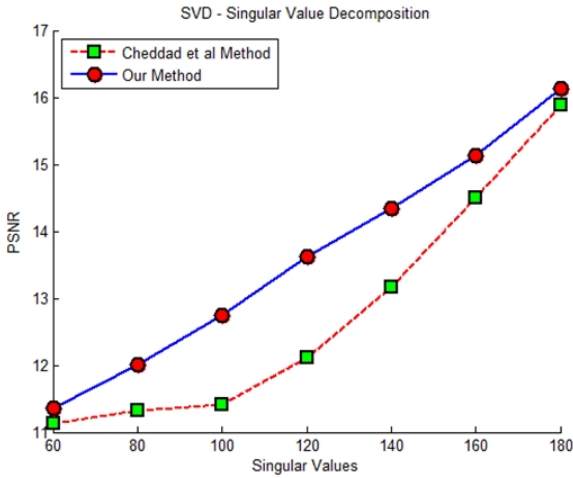


Fig. 6. Results of SVD compression attack on ‘Lena’ image

4.3 SVD Compression Attack

Based on Figure 6, we can see that when 60 and 180 singular values are used for SVD compression attacks, the performances are similar for both methods. When 60 singular values are used, the recovered images from both methods are very noisy. At the same time, when 180 singular values are used, the quality of the recovered images from both methods are more similar. When the singular values between 60 and 180 are used, our method outperformed Cheddad et al’s method.

5 Conclusion

Lossy compression attacks in digital watermarking are one of the major issues when sending digital images over the internet. In this paper, we improved Cheddad et al's self-embedding method to make it resistant to lossy compression attacks such as BTC and SVD. Our experimental results show evidence that the original content of the digital image can be recovered to a certain extent even though the watermarked image is attacked by lossy compression such as JPEG, BTC and SVD. We only experimented with grayscale images. Future work will involve making our method more applicable to a broader range of images, in particular colour images.

References

1. Voyatzis, G., Pitas, I.: Protecting digital-image copyrights: a framework. *IEEE Computer Graphics and Applications* 19, 18–24 (1999)
2. Bender, W., Gruhl, D., Morimoto, N., Lu, A.: Techniques for data hiding. *IBM Systems Journal* 35(3&4), 313–336 (1996)
3. Cox, I.J., Killian, J., Leighton, T., Shammon, T.: Secure spread spectrum for multimedia. *IEEE Transactions on Image Processing* 6(12), 1673–1687 (1997)
4. Vetterli, M., Kovacevic, J.: *Wavelet and Subband Coding*. Prentice-Hall, Englewood Cliffs (1995)
5. Wolfgang, R.B., Podichuk, C.I., Delp, E.J.: Perceptual watermarks for digital images and video. *Proceedings of the IEEE* 87(7), 1108–1126 (1999)
6. Cheddad, A., Condell, J., Curran, K., McKeivitt, P.: A secure and improved self-embedding algorithm to combat digital document forgery. *Signal Processing* 89(12), 2324–2332 (2009)
7. Daren, H., Jiufen, L., Jiwu, H., Hongmei, L.: A DWT-based image watermarking algorithm. In: *IEEE Int. Conf. Multimedia and Expo.*, pp. 429–432 (2001)
8. Joo, S., Suh, Y., Shin, J., Kitkuchi, H.: A New Robust Watermark Embedding into Wavelet DC Components. *ETRI Journal* 24(5), 401–404 (2002)
9. Floyd, R.W., Steinberg, L.: An adaptive algorithm for spatial grayscale. In: *Proc. Soc. Inf. Disp.*, vol. 12, pp. 55–77 (1976)
10. Neelamani, R., Nowak, R., Baraniuk, R.G.: WInHD: wavelet-based inverse halftoning via deconvolution. submitted to *IEEE Trans. Image Process.* for publication (2002)
11. Chanda, B., Dutta Majumder, D.: *Digital Image Processing and Analysis*. Prentice-Hall, Englewood Cliffs (2000)
12. Richards, D., Abrahamsen, A.: *Image compression using singular value decomposition. linear algebra applications* (2001)
13. Prasanth, H.S., Shashidhara, H.L., Balasubramanya Murthy, K.N.: Image Compression using SVD. In: *Proc. of International Conference on Computational Intelligence and Multimedia Applications*, pp. 143–145 (2007)
14. Cheddad, A., Condell, J., Curran, K., McKeivitt, P.: A Hash-based Image Encryption Algorithm. *Opt. Comm. Elsevier Science*. 283(6), 879–893 (2010)
15. Wang, Y., Liao, X., Xiao, D., Wong, K.: One-way hash function construction based on 2D coupled map lattices. *Inf. Sci.* 178(5), 1391–1406 (2008)
16. US Secure Hash Algorithm 1 (2001), <http://www.faqs.org/rfcs/rfc3174>

Generation of View Representation from View Points on Spiral Trajectory

Andrzej Salamończyk¹ and Wojciech Mokrzycki²

¹ Institute of Computer Science, University of Podlasie,
Siedlce, Poland

e-mail: andrzej@ii.ap.siedlce.pl

² Faculty of Mathematics and Informatics, University of Warmia and Mazury,
Olsztyn, Poland

e-mail: mokrzycki@matman.uwm.edu.pl

Summary. In this paper the new method of view representation generation is proposed. View points in the method are located on spiral trajectory enlaced view sphere. Additionally the arrangement of view point is more evenly then in other known similar methods and can be formulated by only one parameter. The number of view points follows from the required scanning resolution.

1 Introduction

Methods of generating 3D multiview representation of polyhedron for object visual identification are described in several papers e.g [1]. Some of them - concerning models for convex polyhedrons only and called *iterative methods* (e.g. presented in [3]), can be described by a series of repeated steps (iteration). Firstly, central views corresponding to object features chosen for identification are generated. Then, *single-view areas* are calculated on the *view sphere*. They correspond to views generated earlier. In each step the covering of the whole viewing sphere by with single-view areas is checked. This process is repeated until we get a complete cover of view sphere.

Papers [6, 7] concern methods called *noniterative*, which are better because they are faster. Complete representation is obtained by covering the viewing sphere precisely with single-view areas without loop, but in spiral way and controlling "edge" register (of not covered area). When the register is set to "empty" the generation of representation is completed. The representation used is complete, which results from the generation method used. However, to achieve complete representation you must calculate single-view areas on viewing sphere and carry them out in a given order. Without single-view areas it is not possible to find a complete set of views. All described methods produce convex polyhedron views only.

Next group of methods [4, 8] does not use any single-view areas, so are faster than the previous ones. They use a concept of *complementary cone* which rotates around faces normals and records every *visual event* to obtain all views. This event occurs as a result of a new normal vector entering the scanning cone or by disappearance of a vector. The outcome of such a routine is a set of vectors faces that can be seen from the view sphere. However, they were tested for convex polyhedrons only [8].

In [9] authors extend this method (the algorithm and its implementation) for monotonous polyhedrons (monotonous polyhedrons are a class of nonconvex polyhedrons including also convex polyhedra). However this implementation do not include shadows in the views. Implementation presented in [10] just includes shadows in the views also. The idea of matching view representations and range images was presented in [12].

For representation generation and obtaining a view we use viewing sphere with perspective projection (*K-M view model* [3]). For the following conditions have to be met:

1. Models are accurate - every model is equivalent to B_{rep} model.
2. Models are viewing models - it is possible to identify object from any view.

We consider polyhedrons that are non transparent and monotonous. As we use a viewing sphere with central projection as a projection space, we allow simple view standardization.

1.1 View Generation Space - Basic Concepts

Let object be a monotonous, non transparent polyhedron without holes or pits. Consider its faces as feature areas, those areas will be used as a foundation for accurate multiview 3D model determining. This model is a set of accurate views, acquired through perspective projection from viewing sphere, according to the model from [3].

The concept of generating 3D multiview representation based on assumed generation space model is as follows:

- Circumscribe a sphere on a polyhedron. The sphere is small (radius r) and its center is at the polyhedron center.
- On this sphere place a space view cone (vc) with angle of flare 2α . This is the *viewing cone*. The vertex of this cone is a model *viewing point* VP . The distance between polyhedron center and model viewing point - R . Viewing axis always goes through sphere (and the same object) center.
- Unconstrained movement of the cone vertex, where the cone is tangent to the small sphere creates a large sphere with radius R . This sphere is called *viewing sphere* (Fig. 1). Each object has its own viewing sphere, the same for all views of this particular object.

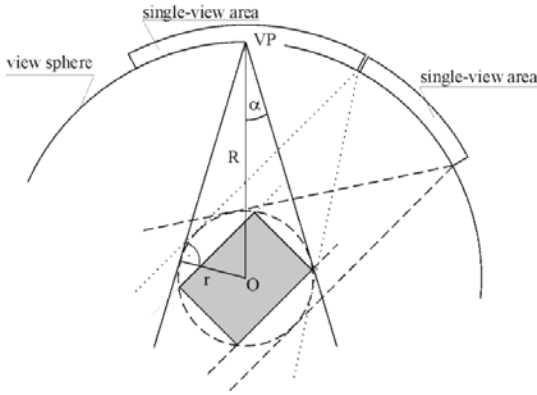


Fig. 1. Concept of view sphere and „single view areas”, [3]

- Generate views, taking into account only object features selected for identification i.e. faces. Faces visible in the viewing cone create a *view*, external edges from this view create *view contour*.

The dependency between r and polyhedron vertices coordinates (X_{vi}, Y_{vi}, Z_{vi}) and: R, α, r and angle of view cone vc flare are:

$$r = \max_{i=1, \dots, k} \sqrt{X_{vi}^2 + Y_{vi}^2 + Z_{vi}^2}, \quad R \geq \frac{r}{\sin \alpha}, \quad \angle(vc) = 2\alpha.$$

Changing one view to the other is a *visual event*. This event occurs as a result of point VP movement and is manifested by appearance of a new feature in a view, disappearance of a feature or both (Fig. 1). *Complementary cone* (cc) is a cone defined by current viewing axis (it's collinear with its height) and has an opposite direction of flare to the viewing cone. It intersects viewing cone with angle $\pi/2$, so its angle of flare is: $\angle(cc) = \pi - 2\alpha$.

View is created by faces that are visible in the viewing cone at a certain viewing point VP position. External edges of a view create a viewing contour. On Fig. 2 the view sphere divided into "one-view" areas is shown. A some view associated with region are also shown.

The other group of method use the idea of obtaining views from view point uniformly distributed on view sphere.

2 Uniformly Distributed View Points on View Sphere

The problem of uniform distribution n points on the sphere has been studied in many branch of science, [5]. Such distribution is useful for generation of view representation (we put view points on the sphere). One of the most popular approach is putting view points in vertices of regular polyhedron (octahedron, tetrahedron, icosahedron) and next we quadrisect of each faces (triangles). Quadrisectation of a triangle consists of inserting a new vertex in

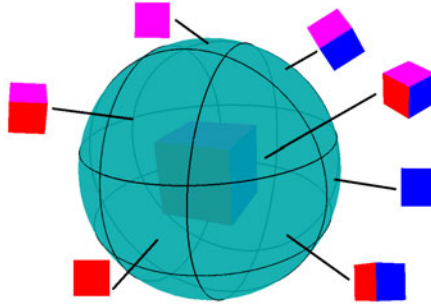


Fig. 2. Some views of a cube and corresponding "one-view" areas on the view sphere

a half of each edge in a such way that original triangle is split into four subtriangles (see Fig. 3). We iterate this process until we get required resolution. Clearly, icosahedron provides the best approximation to the sphere and has been studied in the most cases [2]. Advantages of this method include clear way of obtaining new points (easy to implementation). Nevertheless, this approach has some disadvantages.

Firstly, there is a significant difference between angles (due to the fact that angle $\angle V_{01}V_{12} = \angle V_{12}V_{20} = \angle V_{20}V_{01}$ is grater then angle $\angle V_0V_{01} = \angle V_{01}V_1 = \angle V_1V_{12} = \angle V_{12}V_2 = \angle V_2V_{20} = \angle V_{20}V_0$ see Fig. 3) so the spread of view point is clearly not uniform. The respective values (the most convinient

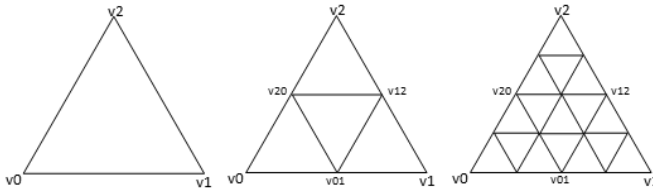


Fig. 3. Quadrisection of a triangle



Fig. 4. Division of a icosahedron in iteration process

Table 1. Differences in minimal angles between points on view sphere in quadrisection of triangle method

Iteration number	Min. angle	Max. angle	Difference
0	1.1071	1.1071	0%
1	0.554	0.628	11.9%
2	0.277	0.326	15.2%
3	0.138	0.164	16.0%

case, the worst case and difference) for angles in successive three iterations of quadrisection of icosahedron are shown in table 1.

Secondly, in each step of iteration the resolution is approximately ~ 0.5 of previous value. So that, in most cases we have to many points to achieve required resolution. For example if we need a cover of view sphere with minimal angle between points equal to 0.5 one subdivision (iteration) is not sufficient (see Table 1) and we have to perform two iterations which ensure subdivision accuracy 0.326.

Last (not least) disadvantage is the fact that the set of view points is not ordered. The process of quadrisection of icosahedron is shown on Fig. 4. View points are located in vertices of generated polyhedron from regular icosahedron.

2.1 View Points on Spiral Path

The idea of spiral scanning of view space on view sphere from which a single face can be visible (and in this way obtaining all views including the considered face) has been shown in [11]. Start point has been coincided with intersection of normal vector of the face and view sphere. Then a spiral scanning (with the constant value between consecutive coil of spiral) was executed. After the spiral motion a full rotation of the complementary cone around face's normal has been performed. During scanning views and visual events have been registered. New views has been added to the database of views of the polyhedron. This process has been repeated for each face of polyhedron and after removing repeated views the complete set of views for a given polyhedron has been produced.

The weakness of a mentioned method which can be easily observed is the fact that some regions on view sphere are scanned many times (as many as number of faces which can be viewed from this region) because the scanning ranges for different faces often overlap (see Fig. 5).

2.2 Uniformly Distributed View Points on Spiral Path

We propose a new method to overcome enumerated difficulties. Our idea is put view points on spiral trajectory in a such way that distribution of points

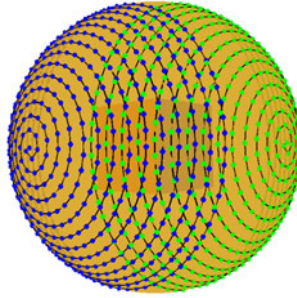


Fig. 5. Overlaped regions in scanning two faces surroundings in spiral way method [11]

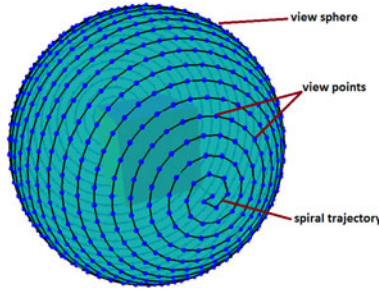


Fig. 6. View points located on spiral trajectory

is uniform (to a certain extent). The formula for spiral trajectory can be defined as follows, [11]:

$$\begin{cases} x = \cos(kt) \sin(t) \\ y = \sin(kt) \sin(t) \\ z = \cos(t) \end{cases} \quad (1)$$

where $t \in [0, \pi]$ denotes a parameter and $2k$ is the number of full rotation around z -axis. However, arrangement of view points determined by the equation (1) is definitely not uniform (more dense at poles and rare on the equator). If we put

$$\begin{cases} t = \arccos(1 - s); \text{ for } s \in [0, 1] \\ t = \pi - \arccos(s - 1); \text{ for } s \in (1, 2] \end{cases} \quad (2)$$

than the points is more uniform distribution. The angle between successive coils of the spiral (the required resolution res) is equal to $res = \frac{2\pi}{k}$ whole length of spiral path defined in (1) is equal to $2k$.

Hence, the formula for the number of view points can be defined as follows:

$$n = \frac{4\pi}{(res)^2} \quad (3)$$

where res is the required scanning resolution. This resolution may depend on technical parameter of device acquired data or complexity of polyhedron. The view points on sample spiral trajectory enlaced view sphere are shown on Fig. 6.

We obtain views from view points using the concept of view sphere with perspective projection introduced in section 1.1.

3 Results

The proposed method has been tested on number of polyhedron and we obtain the same complete set of views as generated by other methods. However we have following advantages:

1. The idea introduced in this paper extends the method of scanning space above each face and due to one scanning we avoid overlaps and the number of view points is significantly less.
2. The number of view point (defined in eq. 3) are determined more accurately (in classical quadrisection of triangle method in each step resolution is approximately 2 times less).
3. The set of view points on view sphere generated by our method is ordered (which can be used for generation of aspect graph).
4. The set of view points on view sphere are more evenly distributed (in comparison to the method of quadrisection of triangle and scanning above each face).

For example, for required scanning resolution 0.2, 97.5% of points differ from each other only by only 3% (the value is the minimal angle to any point on view sphere) (see Fig. 7) so that the view points are more evenly distributed compared to the quadrisection of triangle method (differences can reach $\approx 16\%$ in this case) .

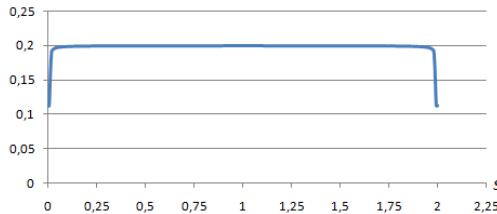


Fig. 7. Distribution of minimal angle to any other point on view sphere depending on parameter value s (eq. 2)

References

1. Arbel, T., Ferrie, F.P.: Informative views and sequential recognition. In: Buxton, B.F., Cipolla, R. (eds.) ECCV 1996. LNCS, vol. 1064, pp. 469–481. Springer, Heidelberg (1996)
2. Belussi, A., et al.: Spatial Data on the Web: Modeling and Management. Springer, Heidelberg (2007)
3. Dabkowska, M., Mokrzycki, W.S.: Multi-view models of convex polyhedron. *MGV* 6(4), 419–450 (1997)
4. Frydler, M., Mokrzycki, W.S.: $2\frac{1}{2}D$ view models of nonconvex polyhedron on view sphere with perspective. In: Proc. Advance Computer Systems 2003, Międzyzdroje, pp. 22–24 (2003)
5. Katanforoush, A., Shahshahani, M.: Distributing Points on the Sphere Experiment. *Math.* 12(2), 199–209 (2003)
6. Kowalczyk, M., Mokrzycki, W.S.: Obtaining complete $2\frac{1}{2}D$ view representation of polyhedron using concept of seedling single-view area. *CV&IU* 91, 208–301 (2003)
7. Kowalczyk, M., Mokrzycki, W.S.: Methods of generation 3D exact views of convex polyhedron for visual identification. Part II: Noniterative methods, implementation and tests results. *MG&V* 12(4), 435–452
8. Frydler, M., Mokrzycki, W.S.: Scanning faces surroundings - new concept in 3D exact multiview of nonconvex polyhedron generation. In: Proc. 4th Int. Conf. on Computer Recognition Systems CORES 2005, Rydzyna, May 22–25, pp. 379–386. Springer, Heidelberg (2005)
9. Mokrzycki, W.S., Salamonczyk, A.: Generating 3D multiview exact polyhedron representation by scanning faces surrounding. *Machine Graphics & Vision* 15(3/4) (2006)
10. Mokrzycki, W.S., Salamonczyk, A.: Using Shaded Views in 3D Multiview Representation of Monotonous Polyhedron. In: IEEE Proceedings 6th International Conference Computer Information Systems and Industrial Management Applications CISIM 2007, June 28–30, Elk (2007)
11. Mokrzycki, W.S., Salamonczyk, A.: Spiral scanning of faces to obtain complete 3D view representation of monotonous polyhedra. In: Computer Recognition Systems 2, Advances in Soft Computing, vol. 45, Springer, Heidelberg (2007)
12. Salamonczyk, A., Mokrzycki, W.S.: New Method of Polyhedron Views and Range Images Matching Using Hausodrf Distance. In: Proceedings of The International Conference on Image Processing & Communications IPC 2009 (2009)

Gradient Based Edge Detection in Various Color Spaces

Marek A. Samko

Faculty of Mathematics and Informatics, University of Warmia and Mazury
e-mail: samko@matman.uwm.edu.pl

Summary. This paper presents gradient based methods of color edge detection in various color spaces. First of all it briefly describes various color spaces and differences between them. Besides it shows edge detection in described spaces. Some examples of applying Sobel's algorithm to real images in various color spaces are presented.

1 Introduction

The most popular color space used in digital images processing is RGB space. This space is used in various devices displaying color images – monitors or displays. That is why this space is commonly used despite its disadvantages. Besides RGB space there is a lot of other color spaces. Each one has its specific features that make it valuable in specific implementations. In this paper four spaces: HSB, YUV, CIEXYZ and CIELab are described. For each of described spaces implementation of Sobel algorithm is shown. All resultant images are presented on white background for better readability.

2 RGB Space

RGB is the most commonly used color space. It was created on the basis of the rule that all visible color could be obtained by mixing red, green and blue color in specified ratios. RGB space is widely used in technical devices (earlier analog, now digital). It is widely used in devices that analyze image (digital cameras, scanners) and devices that display image (monitors, TV sets). That is why it is more popular color space. Unfortunately RGB is theoretical model and its realization could be different in various devices. Every device has its own spectral characteristic for each color channel and could produce quite different range of RGB colors. Thus this space belongs to device dependent color spaces. Besides RGB is not perceptually uniform space. Two colors that are perceived by human as quite different could be placed very close to each other in RGB space. And vice versa, two colors placed in long distance between each other in RGB space could be perceived as very similar.

Sobel's algorithm for image in RGB space is implemented as filtering each color channel separately. Then three edge maps (one for each color channel) are combined in one output edge map using following formula:

$$I(x, y) = \max(R(x, y), G(x, y), B(x, y)) \quad (1)$$

where:

$I(x, y)$ - output edge map,

$R(x, y)$ - edge map obtained after filtering R channel of input image,

$G(x, y)$ - edge map obtained after filtering G channel of input image,

$B(x, y)$ - edge map obtained after filtering B channel of input image.

Fig. 1 shows implementation of Sobel's algorithm for color image in RGB space.



Fig. 1. Lena–input image and implementation of Sobel's algorithm in RGB space

3 HSB Space

HSB (also called HSV) stands for Hue - Saturation - Brightness (Value). This model was created on the basis of human vision manner. It is more intuitive for human when defining colors, because each color is described not by ratios of red, green and blue colors, but by its dye, intensity and brightness. Like RGB this space also is not perceptually uniform.

Sobel's algorithm is applied for each of three channels (hue, saturation and brightness) independently and after that resultant edge map is obtained:

$$I(x, y) = \max(H(x, y), S(x, y), B(x, y)) \quad (2)$$

where:

$I(x, y)$ - output edge map,

$H(x, y)$ - edge map obtained after filtering hue channel of input image,

$S(x, y)$ - edge map obtained after filtering saturation channel of input image,

$B(x, y)$ - edge map obtained after filtering brightness channel of input image.

Fig. 2 shows implementation of Sobel's algorithm for color image in HSB space.



Fig. 2. Lena-input image and implementation of Sobel's algorithm in HSB space

4 YUV Space

The YUV space is typically used in most image and video coding standards [2]. It is commonly used in television signal transmission. Y factor represents luminance of the image, U and V factors represent chrominance of the image. Chrominance components could be efficiently separated from the luminance.

Sobel's algorithm for image using this color space was implemented in following way: Sobel filter was used for filtering luminance and chrominance factors independently. Then output edge map was obtained as maximal response from two filtered channels:

$$I(x, y) = \max(Y(x, y), UV(x, y)) \quad (3)$$

where:

$I(x, y)$ - output edge map,

$Y(x, y)$ - edge map obtained after filtering luminance channel of input image,

$UV(x, y)$ - edge map obtained after filtering chrominance channel of input image.

Fig. 3 shows implementation of Sobel's algorithm for color image in YUV space.

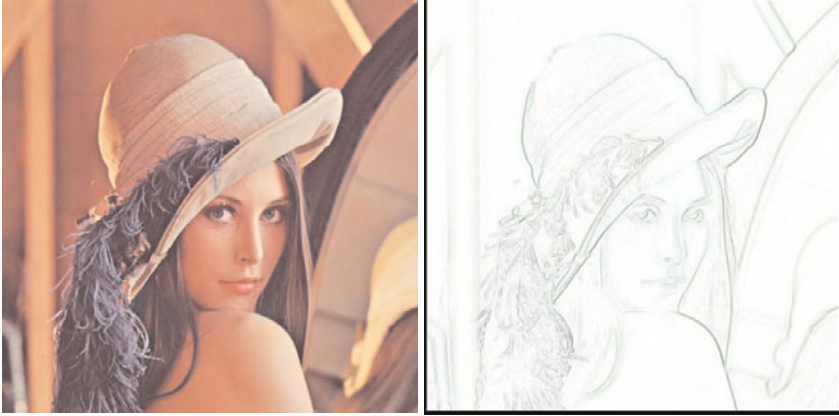


Fig. 3. Lena-input image and implementation of Sobel's algorithm in YUV space

5 CIEXYZ Space

This color space was created to model a way of perceiving color by human. In this space single color is described by three chrominance channels - X, Y and Z. These factors represent ratio of red, green and blue colors in described color. CIEXYZ is also not perceptually uniform space.

Sobel's algorithm for image in RGB space is implemented as filtering each channel (X, Y and Z) separately. Then three edge maps are combined in one output edge map.

$$I(x, y) = \max(X(x, y), Y(x, y), Z(x, y)) \quad (4)$$

where:

$I(x, y)$ - output edge map,

$X(x, y)$ - edge map obtained after filtering X channel of input image,

$Y(x, y)$ - edge map obtained after filtering Y channel of input image,

$Z(x, y)$ - edge map obtained after filtering Z channel of input image.

Fig. 4 shows implementation of Sobel's algorithm for color image in CIEXYZ space.



Fig. 4. Lena-input image and implementation of Sobel's algorithm in CIEXYZ space

6 CIELab Space

This space is mathematical transformation of CIEXYZ space [1]. It was introduced as a result of studies on perceiving differences in colors by human eye. It was assumed that two colors placed in equal distance from each other will be perceived as colors equally different. So CIELab was intended to be perceptually uniform space. Color in this space is described by luminance (L), first color factor (a) - represents color factor from green to magenta and second color factor (b) - represents color factor from blue to yellow.

Sobel's algorithm for image in this space is implemented as filtering each channel (luminance and both color channels) separately. Three edge maps are combined in one output edge map using following formula:

$$I(x, y) = \max(L(x, y), a(x, y), b(x, y)) \quad (5)$$

where:

$I(x, y)$ - output edge map,

$L(x, y)$ - edge map obtained after filtering luminance (L) channel of input image,

$a(x, y)$ - edge map obtained after filtering first color channel (a) of input image,

$b(x, y)$ - edge map obtained after filtering second color channel (b) of input image.

Fig. 5 shows implementation of Sobel's algorithm for color image in CIELab space.

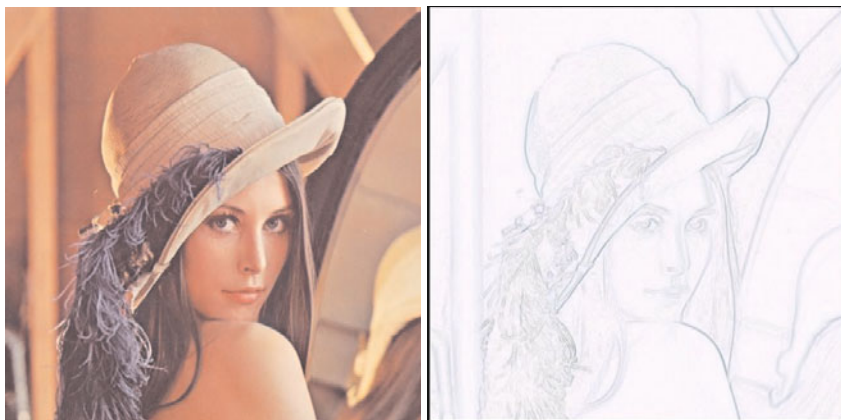


Fig. 5. Lena-input image and implementation of Sobel's algorithm in CIELab space.

7 Conclusions

A brief description of few color spaces and implementation of Sobel's algorithm in these spaces has been presented. In described implementations of the Sobel's algorithm various channels of color image were filtered. Thanks to it we could use information about color (chrominance), not only luminance. Then we have much more information about edges and detection could be more accurate.

References

1. Kuo-Cheng, L.: Color-Edge Detection Based on Discrimination of Noticeable Color Contrasts, Taiwan (2009)
2. Fan, J., Aref, W., Hacid, M., Elmagarmid, A.: An improved automatic isotropic color edge detection technique. *Pattern Recognition Letters* 22 (2001)
3. Malina, W., Ablameyko, S., Pawlak, W.: *Podstawy cyfrowego przetwarzania obrazów*, Warszawa (2002)
4. Samko, M.: Gradient based method of color edges finding. In: Choraś, R.S., Zabłudowski, A. (eds.) *Image Processing & Communications Challenges*, Academy Publishing House EXIT, Warsaw (2009)
5. http://www.knowledgerush.com/kr/encyclopedia/HSB_color_space/ (May 2010)
6. http://www.colourphil.co.uk/lab_lch_colour_space.html (May 2010)

Improve Vector Quantization Strategy

Zahraa F. Muhsen¹, Loay A. Jorj², and Imad H. Alhussaini²

¹ Computer Sciences Department Isra University, Amman, Jordan
e-mail: z_muhsin@ipu.edu.jo

² Computer Sciences Department, Baghdad University, Baghdad, Iraq

Summary. Vector Quantization is an efficient method for image compression. It has been developed as one of the most efficient image coding techniques. It is a process that maps the blocks of high rate digital pixel intensities into a relatively small number of symbols. The aim of this work is to use different ways to encode the homogenous/ heterogeneous or edge/smooth part of the image with the improvement of the existing Vector Quantization algorithms and reduce its complexity. Many techniques in this paper have been examined to improve the quality and the compression ratio for the compressed images, such as the block rotation process, the mean and mode operation, block classification, and random blocks selection. High PSNR results obtain when using scalar quantization as a pre processing with rand selection blocks and blocks rotation.

1 Introduction

The common characteristic of most images is that the neighbouring pixels are correlated and therefore contain redundant information. The foremost task then is to find fewer correlated representations of the image. The two fundamental components of compression are redundancy and irrelevancy reduction [1, 2].

Redundancy reduction aims to remove duplication from the signal. Irrelevancy reduction omits parts of the signal that will not be noticed by the signal receiver, namely the human visual system. Spatial redundancy is a redundancy or correlation between neighbouring pixel values. Spectral redundancy is a redundancy or correlation between different color planes or spectral bands. Any image compression scheme aims to reduce the number of bits needed to represent an image by removing the spatial and spectral redundancies as much as possible [1].

The main objective of data compression is to minimize the average number of bits used to represent a signal in digital form, while maintaining the sufficient data quality. The information flow or storage capacity will approach the maximum rate under the limited transmission bandwidth or storage device.

The reason why data can be compressed is the correlation between data. The main objective of compression technique is to de-correlate the data or remove the redundancy [1, 3, 4].

Vector Quantization (VQ) is a methods that use several signal samples which grouped together to form blocks of data (vectors). VQ is used to represent an image by decomposing it into collections of blocks known as reproduction vectors; the collection of these vectors is called codebook. A codebook is generated from a set of finite training vectors for representing the input vectors. VQ requires a large amount of computation. The most time consuming factor in VQ is the "nearest neighbour search"; this search is followed to find the vector nearest to an input data vector among a large number of reference vectors (codebook) [5, 6, 7, 8].

2 Adaptive VQ-Design

Large vectors (blocks) will lead the VQ encoder to become complex, because this will require many computational resources (e.g, memory, computation per pixel) in order to efficiently construct and search a codebook. One of the main drawbacks of the VQ compression scheme is the visible blocking distortion that may appear in the reconstructed images. This is due to the fact that encoded images are dissimilar to those present in the training set [9]. New ideas have been introduced to solve the above problems such as, the block rotation process which gives encouraging result toward constructing near-optimal codebook. The **Mean** and **Mode** are two important statistical parameters, which used with variance to assess the homogeneity of the block. Removing the blocks that has one color will reduce the number of training blocks, and a pre processing step which determine the homogenous/heterogeneous of the training blocks could also help to reduce the search time. Choosing random blocks is a simplest conceptual approach for initializing the codebook with N reproduction vectors, in which a good codebook may generate with this method.

2.1 Rotation Block

Each tested vector is rotated and reflected (to produce 8 versions), then each version is tested with the codebook contents. When one of the 8 blocks is the closest to one of the codebook vectors. Fig. 11 visualizes the rotated blocks at different θ in a "clown" image. Fig. 11a, 11b and 11c represent the rotation at $\theta=90, 180$, and 270 . Image on Fig. 11d reflects at X-axis's. In Fig. 11e, 11f and 11g represent the rotation of the reflection at X-axis's with $\theta=90, 180$, and 270 . In Fig. 11 black block represent the rotation block match with other black box.

2.2 Mean and Mode

Mean is the average value of the block, **Mode** represent the value that got high redundancy in the block. Fig. 12 shows an example of images where blocks

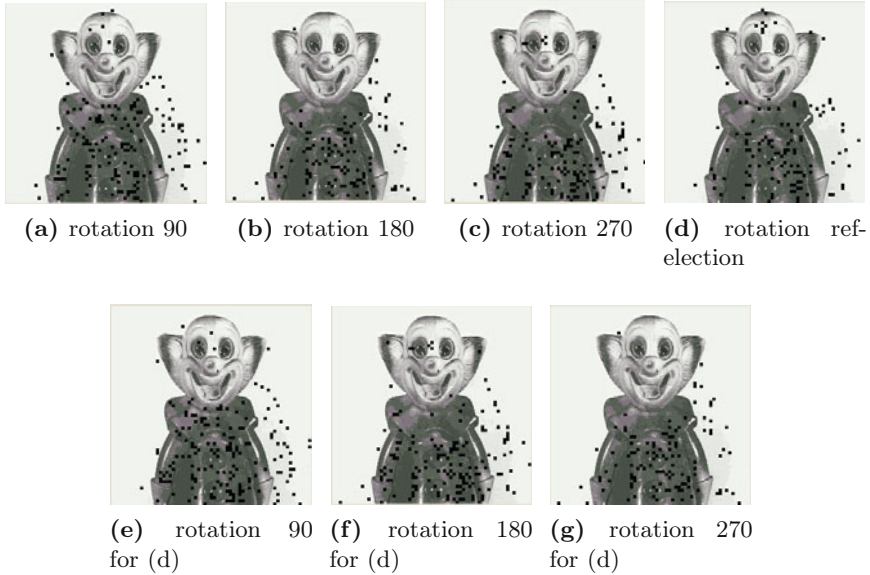


Fig. 1. Matching Results with Rotated-Reflected Blocks

represented in terms of its **mean**, **mode**, or choosing a values between them. In Fig. 2f some artificial edges and smoothness can be seen. The difference in the tiger’s eyes clearly shows the difference between using the **mode** or **mean** values. **Mean/mode** has been used as criteria for all the values of the blocks. They used to assess the contrast among the values of the block. The disadvantage of using the **mean** value is: if there is a large difference between the values of pixel, **mean** got a value not close to any of the pixels values. While in **mode**: when most of the pixels values have the same redundancy, then none of them represent a good indicator for the block contents. When using both **mean/mode**: if the redundancy is larger than the half of the dimension of the block then **mode** is a good choice, else the **mean** will be better, Fig. 2d illustrates this option.

The example in Fig. 3e and Fig. 3f shows the **mean** and the **mode** of a given block. The **mode** is better to represent the block’s content; the **mean** is useful when all the values of the block have the same redundancy.

2.3 Mean Removed VQ (MRVQ)

In MRVQ method the blocks that represented by the mean/mode of the training set vectors are removed before the VQ process. Vectors with a lot of energy will have great distance from the origin, while vectors with small amounts of energy will situated quite near the origin. This method is very similar to the pruning method 7 in that the initial codebook is created from vectors taken directly from the training set.

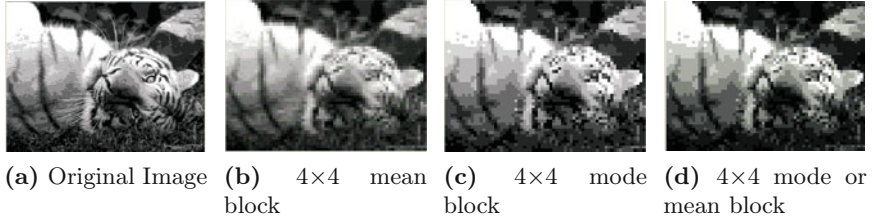


Fig. 2. Examples of Mean and Mode Block 4×4

2.4 Block Classification

First the image is partitioning into homogenous and heterogeneous blocks. The homogenous could have different smoothness when input to VQ. Some blocks may have single tone while others may have edge information. Some blocks have a big dispersion between their elements other have less. Classifying blocks into different categories could reduce the search time. This classification is always coming before the process that deals with codebook generation and blocks matching. The blocks have been classified as follow:

- **Smoothed block:** when the variance of the block is less than Ω_{edge} .
- **Edge block:** when the variance of the block is larger than or equal to Ω_{edge}

Each set of classified blocks deals with the design or the clustering process in different way. The *smoothed blocks* are considered as their contents consist of single gray-level, and the value of this gray level is set to either the mode or the mean. Fig. 3b and Fig. 3c shows different *smooth block*.

Fig. 3a is the original image. The smoothed blocks (variance $< \Omega_{edge}$) replaced by black block. Fig. 3d, 3e and 3f, shows the edge block whose variance greater than Ω_{edge} .

2.5 Random Coding

This method generates the codebook by using random generator based on the statistical distribution of the training sequence. A good codebook may generate by using this method but it is more prone to fall into local optimum's than other complex methods. The initial codebook can be filled simply by randomly choosing vectors from the training set image block. This can be started by taking the first random vector from the training set (i.e choose randomly find x,y for 2D, the index of image block for 1D), and place it into the codebook. The distortion between the randomly selected vectors and the training set is calculated. If the distortion is greater than a predetermined threshold (Ω_{dis}), then the vector is added to the codebook. The process is continued until the codebook is filled or all the blocks of the training set are checked.

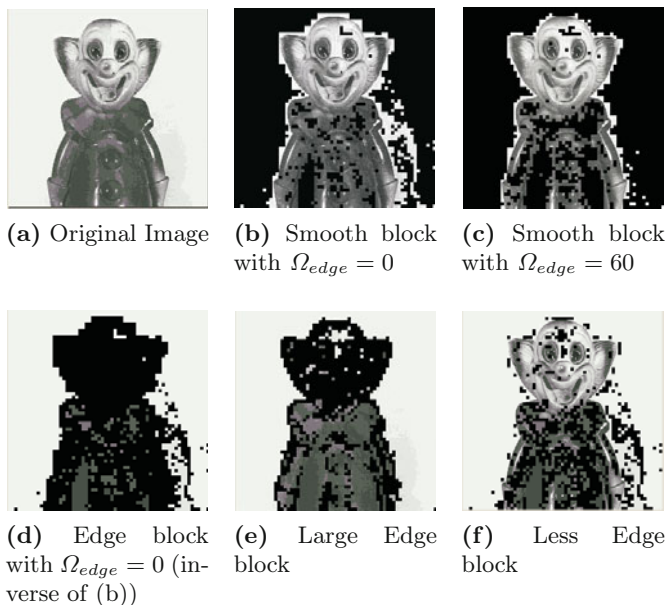


Fig. 3. Types of Classified Blocks

2.6 Pruning Found

It is a very common method of creating an initial codebook from the training vectors. It started with the training vectors and selectively eliminating training vectors as candidate code vectors until a final set of training vectors remains as the codebook. Usually, this involves computing the block difference between each training set vector (with N vectors in the training set) to find its adjacent training set vector. The vector pair with closeness replaced with the centroid of the pair, thus leaving N-1 vectors. This is implemented recursively until the required number of vectors is left.

A modified method of pruning method is introduced, which mapped point techniques will used. This map represents the training block, if the training block is eliminated by pruning methods then set its value map point to 0, otherwise set it will be 1. The following steps are the algorithm of the **Map point**:

Fig. 4 shows the pruning map point process. In this figure the "one (1)" means this block have been checked with all blocks and it is either the centroid of the cluster or it does not find any block near to it by Ω "Zeroes (0)" block means these block is out of check and does not considered as a centroid of an cluster.

This search algorithm is fast search, since the blocks are checked sequentially for one time only, and in this way the search time is less than that required by other methods. In this technique the initialized codebook block are the most different blocks from each other, so a pruning method could

Algorithm 2. Map point

1. Initialize the value of the map point of all the blocks of the training set with 1. The Map cell represented one bit.
 2. **for** for all blocks listed in the training table **do**
 3. Read a block from the training set. The map point corresponds to this block must be 1.
 4. Check sequentially each training block with all other training blocks with map point is 1, but not with itself.
 5. **if** the checked block is very close to other training block **then**
 6. set the map point of all closed blocks with 0.
 7. **end if**
 8. **if** there is no close block to the examined block **then**
 9. do not change its corresponding map point
 10. **end if**
 11. **end for**
 12. Set the training blocks whose map points equal to 1 as codebook vectors.
 13. **if** the size of the codebook exceeded the required number **then**
 14. eliminate the codebook vector by removing the vector whose population is low.
 15. **end if**
-

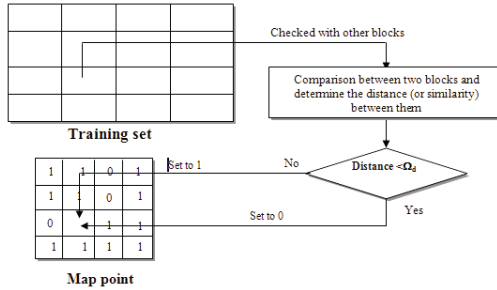


Fig. 4. Map point Process

not be applied further. That, the codebook will consist of two types of vectors: redundant vectors the most redundancy vector and the most different ones, which will depend on the value of Ω to define the degree of the nearest neighbour methods.

The modification of pruning method leads to less complexity in computing and computation time. In this modification the step one is more efficient for generating. The improvement of this method is based on the fact that using a *flag mapped* (formed with N cells), each cell may have the value **0** or **1** (i.e. the training block had been clustered or not).

2.7 Result and Conclusion

Rotation, remove mean/mode, and SQ have been added to the traditional VQ techniques (pruning and random), Table 1 shows results of applying the modification process.

Table 1. Results of the Rand VQ Techniques on "Lena" Image with different modification

N	Methods	Book size	BitR	PSNR	Time Sec
1	Rand VQ Techniques	128	0.687	26.203	19
		64	0.500	26.168	20
		32	0.375	25.901	21
2	Rand With Rotation Process	128	0.875	26.240	724
		64	0.687	26.186	321
		32	0.562	26.203	708
3	Rand WithRotation & Remove Mean/Mode process	128	0.939	19.482	70
		64	0.826	19.246	62
		32	0.760	19.243	42
5	Rand With Rotation, Remove Mean/Mode & SQ	128	0.934	48.688	23
		64	0.822	47.182	12
		32	0.756	47.014	10

As observed from the result of Table 1, combining rand with rotation remove mean /mode and SQ process exhibits the best result comparing to the others in both PSNR and Time sec, the values shown in boldface represent the best result for the processes.

The important parameters that effect the execution of the methods used by the VQ techniques are the edge/smooth blocks (Ω_{edge}), and the distance (Ω_{dis}). The values of these two parameters could be adjusted (in an iterative way) to reach the better performance.

The Distance Ω_{dis} parameter affect on the initialization of the codebook. It is so important for determining the search time. When the value of Ω_{dis} parameter is big the number of codewords is small and will affect the finding of the missing codewords. This will affect the attributes of the tested methods since most of the codeword will be found by the random way.

3 Conclusion

In our previous discussions we have reported some remarks related to the behaviour and performance of the suggested coding processes. A summary of some important conclusions could be as follow,

- The selection of parameter values should be chosen correctly, because these values have a grates effect on the compressed image quality.
- The map point process is very fast.

- Experimental results indicate that the use of SQ combined with VQ give a uniform values.
- The remove of the mean/mode has provided the codebook a generality for all other test image.
- The pre processing on the VQ gives it the reduction in speed and complexity.
- Assigning a few bits to likely symbols (smooth area) and more bits to unlikely symbol (edge area) gives a fine vision to the resulted image.
- The selection of the quantization methods should be chosen carefully, because experimental results show that each one have a view point in one of the measurement.

References

1. Salomon, D.: A Concise Introduction to Data Compression. Springer, London (2008)
2. Gonzales, R.C., Wintz, P.: Digital Image Processing. Prentice-Hall Inc., Upper Saddle River (2008)
3. Cazuguel, G., Cziho, A., Solaiman, B., Roux, C.: Medical Image Compression and Feature Extraction using Vector Quantization, Self-organizing Maps and Quadtree Decomposition. In: Information Technology Applications in Biomedicine (ITAB 1998), Washington DC, May 16-17 (1998)
4. Hong, E.S.: Group Testing for Image Compression. PhD., University of Washington, Computer Science and Engineering (2001)
5. Gray, R.M., Neuhoff, D.L.: Quantization. IEEE Trans. on Infor. Theory 44(6), 2325–2384 (1998)
6. Marcellin, M.W., Lepley, M.A., Bilgin, A., Flohr, T.J., Chinen, T.T., Kasner, J.H.: An Overview of Quantization in JPEG 2000. Signal Processing Image Communication 17, 73–84 (2002)
7. Cosman, P.C., Oehler, K.L., Riskin, E.A., Gray, R.M.: Using Vector Quantization for Image Processing. Proc. IEEE 81(9), 1326–1341 (1993)
8. Fisher, Y.: Fractal Image Compression Theory and Application. Springer, New York (1994)
9. Cosman, P.C., Gray, R.M., Vetterli, M.: Vector Quantization of Image Subbands: A Survey. IEEE Trans. On Image Processing 5(2), 202–225 (1996)

Biometric

Knuckle Biometrics for Human Identification

Michał Choraś and Rafał Kozik

Image Processing Group,
Institute of Telecommunications, UT&LS Bydgoszcz
e-mail: chorasm@utp.edu.pl

Summary. In this paper we present human identification method based on knuckle biometrics also termed as FKP (finger-knuckle-print). Knuckle is a part of hand, and therefore, is easily accessible, invariant to emotions and other behavioral aspects (e.g. tiredness) and most importantly is rich in texture features which usually are very distinctive. The major contribution of this paper are texture-based knuckle features and their evaluation using IIT Delhi knuckle image database.

1 Introduction

Even though biometric identification systems became our reality and are no longer science-fiction visions, only several modalities have been widely deployed and such systems still have many drawbacks. The most known and often used modalities are fingerprints, face, hand geometry and iris. These are widely deployed in large-scale systems such as border control and biometric passports. But due to the problems with large-scale scalability, security, effectiveness and last but not least user-friendliness and social acceptance (e.g. even some governments protested vs. American policy of fingerprint enrollment for visitors) new emerging modalities are still needed.

Therefore, in this paper we present our approach to identify humans on the basis of their knuckles. Knuckle is a part of hand, and therefore, is easily accessible, invariant to emotions and other behavioral aspects (e.g. tiredness) and most importantly is rich in texture features which usually are very distinctive. Knuckle biometrics methods can be used in biometric systems for user-centric, contactless and unrestricted access control e.g. for medium-security access control or verification systems dedicated for mobile devices (e.g. smartphones and mobile telecommunication services).

Even though knuckle biometrics is relatively unknown and new modality, there already are some results and feature extraction methods. So called KnuckleCodes have been proposed and other well known feature extraction methods such as DCT, PCA, LDA, ICA, orientation phase, Gabor filters have been investigated showing very good identification results [\[1\]](#) [\[2\]](#) [\[3\]](#) [\[4\]](#) [\[5\]](#).

Hereby, we propose texture feature extraction methods such as Probabilistic Hough Transform (PHT) and Speeded Up Robust Features (SURF) and the original 3-step classification methodology.

The paper is structured as follows: in Section 2 general view on our system architecture is presented. In Section 3 knuckle image preprocessing is described. Feature extraction methods are proposed in Section 4. Classification methodology is presented in Section 5. Results and conclusions are given thereafter.

2 Knuckle Biometrics System Architecture Overview

The architecture of our knuckle biometrics system is presented in Fig 1.

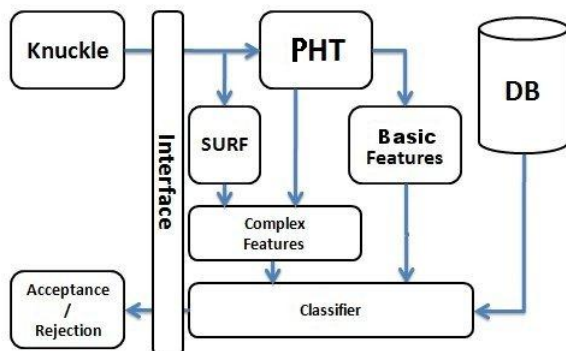


Fig. 1. Architecture of the proposed knuckle biometrics system

Firstly, the knuckle image is obtained from individual requesting access to the system. The knuckle image is preliminary processed to gain the characteristic features. The preprocessing includes both edge detection and thresholding. The image is further analyzed by means of Probabilistic Hough Transform (PHT), which is used both for determining the dominant orientation and also for building the basic feature vector. We also calculate enhanced feature vector using PHT output giving the input for final classifier which uses the SURF texture features.

Then the 3-step classification methodology is applied (in a broad-narrow manner). For computed "basic feature vector" nearest neighbors yielding the shortest Euclidean distance are chosen. For each image in kNN set the complex feature vectors are compared. The approach with kNN allows to decrease the complex computation without losing the overall system effectiveness (as discussed in details in section 6).

3 Knuckle Preprocessing Phase

The most noticeable knuckle texture features are the lines and wrinkles located on bending area of finger joints (see the first row in Fig. 3).

Therefore our methodology focuses on extracting these lines. First the image is binarized using an adaptive threshold estimated by means of equation 1:

$$T = \mu - \frac{\delta}{6}, \quad (1)$$

where T indicates the threshold value, μ the mean value and δ the standard deviation. Both the mean value and the standard deviation are computed locally in blocks of 7×7 pixels size.

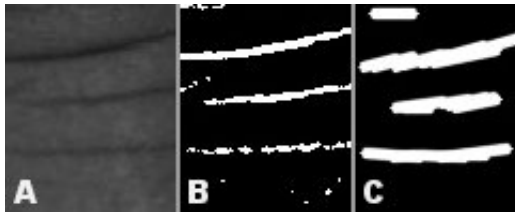


Fig. 2. The knuckle image example (A), enhanced major lines after thresholding (B), and the lines detected by PHT (C)

The result of adaptive thresholding can be seen in Fig. 2 B. It can be noticed that such an image is quite noisy, since some edges suffer from line discontinuities while the background is filled with small spots. This problem is solved by adapting the Probabilistic Hough Transform (PHT). Later, the PHT is also used in our approach for extracting the dominant orientation and building the "basic feature vector".

4 Feature Extraction

4.1 Short Feature Vector (Basic Features)

The basic features vector describing the knuckle texture is built using the PHT output information, which contains a set of line descriptors represented by formula 2, where $LD_i(N)$ stands for N -th line descriptor of i -th image, (b_x, b_y) the Cartesian coordinates of line starting point, (e_x, e_y) the Cartesian coordinates of line end point, θ the angle between the line normal and the X-axis, and d the particular line length expressed in pixels.

The number of extracted lines (N) depends strictly on knuckle spatial properties and varies, therefore these are not directly used to build feature vector.

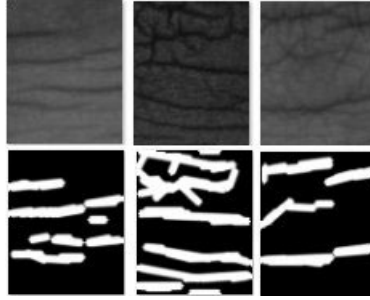


Fig. 3. Sample knuckle images and their representation after applying PHT transform

$$LD_i(N) = [b_{xN}, b_{yN}, e_{xN}, e_{yN}, \theta_N, d_N] \quad (2)$$

Due to the fact the particular knuckle may be rotated, the dominant orientation based on Hough transform is extracted using the θ angle from the line descriptors, which is used to rotate analyzed image in opposite direction to align the dominant line perpendicular to y-axis. After that the y position of particular line and its length is used to build the feature vector. The 30-bins 1D histogram is adapted.

The vectors described in this section were named "basic" since these are relatively short (one row vector of length 30) and are used for general data set clustering to decrease the number of computations and comparisons of complex features vector in further phases of our human identification system.

4.2 Knuckle Lines Model

The set of line descriptors (eq. 2) obtained from Hough transform are converted to image representation giving input for matching algorithm. Both query and template images (chosen from kNN selected by basic feature classifier) are transformed and compared using the Euclidean metric. The output of matching block is the scoring map, which is 30x30 of size. The size is determined by searching ranges. In this case the template image is offset in $\langle -15, 15 \rangle$ range both on x and y dimension as is it defined by formula 3, where i and j are defines the element in scoring map, H and W defines query image width and height respectively, q and t represents query and template images respectively.

$$score(i, j) = \sum_{x=0}^W \sum_{y=0}^H (q(x, y) - t(x + i, y + j))^2 \quad (3)$$

The lowest score (the shortest distance) is selected giving the information about how the query image is similar to template, and allows to handle offsets

in knuckle images. This is necessary since the knuckle database images are acquired using a peg-free method [6].

The 5 images from kNN set yielding the lowest score are chosen giving the input for SURF-based classifier.

4.3 Knuckle Texture Descriptors

The SURF stands for Speeded Up Robust Features and is robust image detector and descriptor. It was firstly presented by Herbert Bay in 2006 [9]. It is widely used in object recognition and 3D reconstruction. The key-point of the SURF detector is the determinant of the Hessian matrix, which is the matrix (eq. 4) of partial derivatives of the luminance function.

$$\nabla^2 f(x, y) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \quad (4)$$

$$\det(\nabla^2 f(x, y)) = \frac{\partial^2 f}{\partial x^2} \frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 \quad (5)$$

The value of the determinant (eq. 5) is used to classify the maxima or minima of the luminance function (second order derivative test). In the case of SURF the partial derivatives are calculated by convolution with the second order scale normalized Gaussian kernel. To make the convolution operation more efficient the Haar-like function are used.

If the determinant value is greater than threshold (estimated during experiments on learning data set) then it is considered as a fiducial point. The greater the threshold is the less points (but strong ones) are detected. For each of the fiducial points the texture descriptor is calculated.

In our approach we use the SURF points to find the closest matching (if any) between querying image and the templates selected by PHT-based classifier. Firstly the points yielding the Hessian determinant value greater than threshold are selected for both querying and the template images resulting in two points data set. Basing on texture descriptors the matching pairs between those sets are found and the outliers (points in one data set that do not

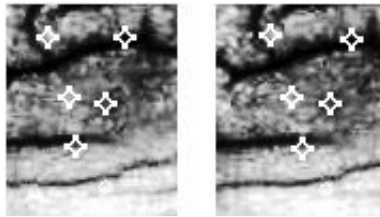


Fig. 4. Detected fiducial SURF points for querying image and its corresponding matches for the template image

have representative in the second data set) are removed. Then the matching cost between those sets is estimated using eq. [6](#):

$$m_{cost} = \sum_{i=0}^N d(p_i - \frac{1}{N} \sum_{j=0}^N p_j, q_i - \frac{1}{N} \sum_{j=0}^N q_j), \quad (6)$$

where N , d , p and q represents the number of matching pairs, Euclidean distance, point from template image and point from query image respectively. Example of such a mapping is shown in Fig. [4](#).

5 Classification

Hereby we propose classification methodology that consists of 3 consecutive steps: selecting 50 images on the basis of basic vector, then selecting 5 images on the basis of PHT feature vector, and finally SURF feature vector is used.

When basic feature vector for particular knuckle image is computed it is looked up in data base to find k nearest neighbors yielding the nearest Euclidean distance. The k number was determined empirically as an compromise of system effectiveness and system performance. Classification error is decreasing significantly when the number of neighbors (k) is increased. Basing on experiments we set these number to 50.

For each object form k nearest neighbors the PHT-based method is used to obtain 5 closest matching. For each of these images only one is chosen. In case the SURF-based classifier fails and is unable to find matching template then the first nearest neighbor obtained from PHT is returned with appropriate matching score.

6 Results

The proposed approach was tested using IIT Delhi Knuckle Database [6](#). The knuckle images were obtained from 158 individuals. Each individual contributed five image samples which implies 790 images in database. The database was fully acquired over a period of 11 months.

For efficiency assessment the 5-fold method was applied (the same method as the authors of the database applied in [4](#)) and average of experiments results is presented. The average equal error rate obtained during experiments is 1.02%.

The table shown in Fig. [6](#) shows the EER deviation from its mean value and EER during each of experiments. The FAR and FRR vs system threshold for one of the experiments is shown in Fig. [5](#).

The experiments show that combination of PHT and SURF gives better results than each of this method used separately. The PHT gave 95.56% classification error while the SURF 85.75%. The SURF failed so often due to the fact it was unable to find matching between query knuckle image and the

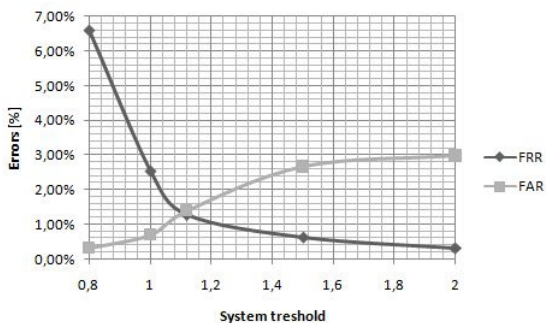


Fig. 5. FAR vs. FRR

Experiment	#1	#2	#3	#4	#5	AVG. EER
EER	0.32%	1.30%	0.95%	1.26%	1.26%	1.02%
					STD:	0.50%

Fig. 6. ERR obtained during experiments

template. Those fails were covered by PHT. However the PHT failed when it came to distinct two or more similar knuckles in k nearest neighbors. In this situation SURF was more accurate.

Obtained results suggest that using simple and fast line and texture extraction techniques is promising and may yield satisfactory results.

7 Conclusions

In this paper new developments in human identification based on knuckle texture features are presented. The major contribution of the paper are: new knuckle feature extraction methodology based on PHT (Probabilistic Hough Transform) and SURF features as well as original classification methodology. The reported results are very good and comparable (slightly better) to other methods (e.g. [4]). However, our methods are efficient and fast so that can be applied to contactless biometrics using mobile devices in the very near future [10].

References

1. Morales, A., Ferrer, M.A., Travieso, C.M., Alonso, J.B.: A knuckles texture verification method in a transformed domain. In: Proc. of 1st Spanish Workshop on Biometrics (on CD), Girona, Spain (2007)
2. Kumar, A., Zhou, Y.: Human Identification using Knuckle Codes. In: Proc. BTAS (2009)

3. Kumar, A., Ravikanth, C.: Personal authentication using finger knuckle surface. *IEEE Trans. Information Forensics and Security* 4(1), 98–110 (2009)
4. Kumar, A., Zhou, Y.: Personal identification using finger knuckle orientation features. *Electronics Letters* 45(20) (2009)
5. Zhang, L., Zhang, L., Zhang, D.: Finger-knuckle-print verification based on band-limited phase-only correlation. In: *Proceedings of the 13th International Conference on Computer Analysis of Images and Patterns*, pp. 141–148 (2009)
6. http://www4.comp.polyu.edu.hk/~csajaykr/myhome/database_request/knuckle/
7. Zhang, L., Zhang, L., Zhang, D., Hailong, Z.H.: Online Finger-Knuckle-Print Verification for Personal Authentication. *Pattern Recognition* 43(7), 2560–2571 (2010)
8. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. *Computer Vision and Pattern Recognition*. In: *Proceedings of the 2001 IEEE Computer Society Conference on CVPR 2001*, vol. 1, pp. 511–518 (2001)
9. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
10. Choraś, M., Kozik, R.: Feature Extraction Method for Contactless Palmprint Biometrics. In: Huang, D.-S., et al. (eds.) *ICIC 2010*. CCIS, vol. 93, pp. 435–442. Springer, Heidelberg (2010)

A New Method of Fingerprint Key Protection of Grid Credential

Yarema Varetsky¹, Bogdan Rusyn¹,
Agnieszka Molga², and Anatoliy Ignatovych³

¹ Karpenko Physical-Mechanical Institute Of NAS of Ukraine
e-mail: jvaretsky@ipm.lviv.ua, rusyn@ipm.lviv.ua

² Kazimierz Pułaski Technical University of Radom
e-mail: agnieszka19216@wp.pl

³ Lviv Polytechnic National University
e-mail: ignatovicha@gmail.com

Summary. In this paper a novel method of biometric protection of private keys is consider. This method is based on enhanced Juels and Sudan "fuzzy vault" scheme. Introduced algorithms utilize minutiae based fingerprint data for key locking procedures. It is proposed to include derived cryptographic primitive into special «biometric» extension of X.509.v3 certificates, which are used in grid environment authentication procedures.

1 Introduction

In paper [1] the possibility of biometric data embedding in x.509 certificates for grid systems is consider. It is very perspective area of cooperation of biometric identification technology and grid-computing, because X.509.v3 certificates are used to encrypt and sign documents and it is impossible for now to tamper with certificates, but there is no mechanism apart from biometrics that can ensure that a certificate belongs to the same person who claims to be an owner of the certificate.

Now we propose an approach to biometric protection of private keys. Our method is based on enhanced Juels and Sudan "fuzzy vault" scheme [2]. We refine this scheme by using minutiae based fingerprint data.

2 Method of Fingerprint Protection of Private Keys

Initially it is necessary to create some cryptographic primitive, which bond private keys from user grid credential with fingerprint's minutiae. Formally we must create some set of data, which lock in some way our private keys.

This set creation occurs during new user registration process or existing user key changing process.

Proposed algorithm creates from N input sets of minutiae locking set

$$W = \{w_0, \dots, w_s\} \subset F$$

of private cryptographic key

$$M = \{m_0, \dots, m_{k-1}\} \subset F$$

which is written in the form of polynomial coefficients $f(x)$ degree $k - 1$ over the field

$$F : f(x) = m_1 + m_2x + \dots + m_kx^{k-1}.$$

To achieve cryptographic resistance of locking algorithm it is necessarily to add to the set of W a number of dummy minutiae. Total number of minutiae after the addition is r . It is found out, that r determined by the minimum possible value of the distance L between minutiae, which is strictly bigger than $2\sigma_s$, where σ_s - average distance threshold, which depend on technology of obtaining minutiae sets (scanner specifications, properties of methods for image processing, etc.) That is, the less L , the more value of r , as follows, more resistant cryptographic protection. But, experimentally determined, that with increasing r increases the probability of false decoding, or false reject rate (FRR). On the other hand the number and location of dummies minutiae are limited to real deployment of personal minutiae and standard deviation of the placement of these minutiae.

The result of the locking algorithm is shown in Fig.1.

Let's define the parameters of the algorithm k, s, r . The polynomial coefficients $f(x)$ are elements of some finite field $F = GF(n)$. Since in the proposed approach we solve the Polynomial Interpolation problem then "fuzzy vault" scheme offers the use of the Reed-Solomon error correction codes (RSC). For the performance of the algorithm for decoding of the code should use field of $n = g^z$ elements, where g - prime number. At the same time the field for locking algorithm is a pixels set of fingerprint image. In our experiments it was used scanners with scan area 322×255 pixels. The efficient scanning area is a region of 280×249 pixels., which accounts for the main part of the image ($\approx 98\%$). This conclusion gives a choice of a simple number. The nearest to above values is a region of an area 251×251 pixels, that is forming field $F = GF(251^2)$.

The polynomial length, of k value, is determined by length of locked private key. Each element of the field $GF(251^2)$ brings 16 bit of information. That is for 256 bit private key of length correspond 16 polynomial coefficients, or polynomial of 15 degree.

To convert a coordinate position of real and fictitious minutiae in field's elements it is appropriate to use 16 bit integers x_i , which 8 lower bits corresponding to ordinate, and high-order bits - to abscissa.

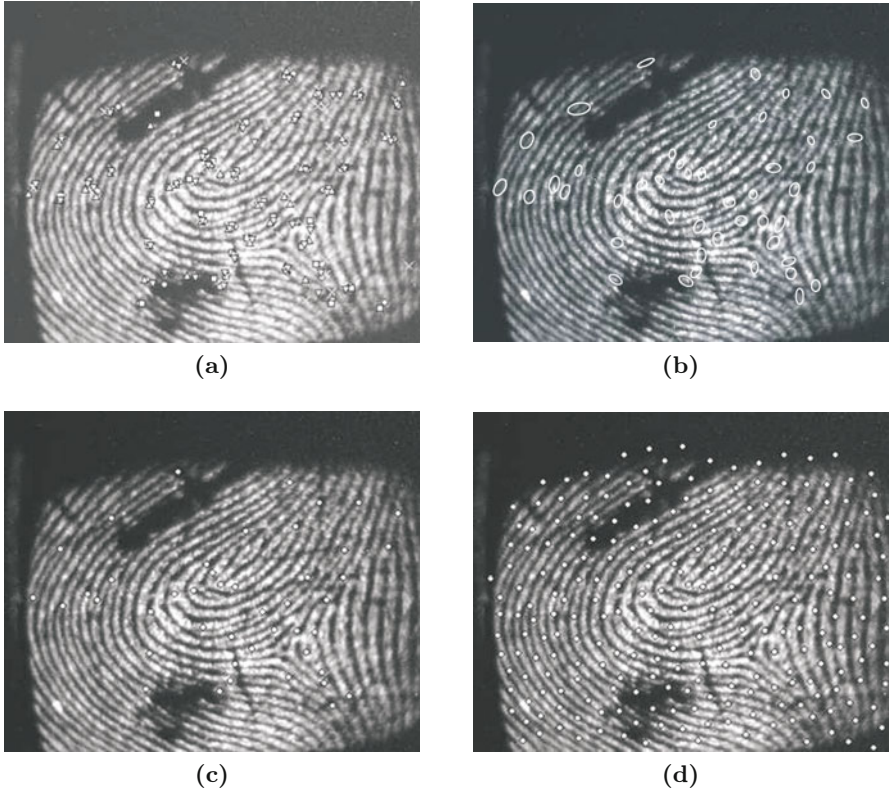


Fig. 1. Creation of locking set. (a) $N = 4$ minutia's sets (symbols $\circ, \square, \times, \triangle$ -minutiae from one set); (b) the most possible sites of user's minutiae locations; (c) minutiae sites centers of gravity - generated locking set; (d) locking set with dummy minutiae.

At the exit of the lock algorithm obtained a set of tuples B_P , consisting of s pairs $\{w_i, f(x_i)\}$ and $r-s$ pairs of dummies points $\{\alpha_i, \beta_i\}$ of F , that satisfy the conditions $f(\alpha_i) \neq \beta_i$. To open system that uses B_P , attacker should extract this set of points, lying on polynomials $f(x)$, thus recover the private key. Evidently, that the larger is r the greater is the number of similar to $f(x)$ false polynomials so, greater is the resistance of the system for exposure. For legitimate system user it is required and sufficiently to present at least $\tau \geq k$ real points, to successfully interpolate implicit polynomial. For unlocking algorithm it is given a set $W' \subset F$, which contains only part of elements W , that is difference of two sets of minutiae is equal to $\#(W - W') = t$.

To unlock the encryption key from BP user provides a set of personal minutiae, forming a unlocking set $W' = \{w'_1, \dots, w'_r\}$. Unlocking procedure takes place when user ack system for protected key issue.



Fig. 2. Locking set \circ with superpose unlocking set \square . Real minutiae are highlighted (radius $\sigma_S = 6$)

Via W' and B_P the set B'_P of nearest (threshold distance σ_S) minutiae with power r is extracted, where $r \approx s$ for legitimate user - $r \gg s$ - for illegal user. The result of minutiae extraction process is shown on Fig. 2. This figure signs - $r = 47$, code word length - $s = 45$, minutiae match - $\tau = 35$. For value $k = 16$ minimum possible number $\tau = \frac{s+k}{2} = \frac{45+16}{2} = 31$.

3 Method Verification

Using a similar approach, it is easy to evaluate the unlock algorithm efficiency depending upon the number of true minutiae.

Notice: the greater is r , the more difficult is unlocking process for illegal user, but also increases the complexity for the legitimate user (Fig. 3, $k = 16$, $s = 45$).

Another important parameter that affects the efficiency of the algorithm is k . Fig. 4 illustrates the complexity of the attack as a function of k and the number of real points.

Dependences show decreasing resistance and increasing FRR with increasing correction capabilities of biometrics uncertainty.

It is proposed to include the resulting set B_P in to special «biometric» extension of X.509.v3 certificate. According to the changes, delegation of user rights in grid environment performs the following procedures: after mutual authentication of user and service, who will work on behalf of the user, service creates a new key pair and sends the public key to the user for signature; user signs the key, similar to, the center of certification, using blocked in B_P private key. Received proxy-certificate and a newly generated temporary key

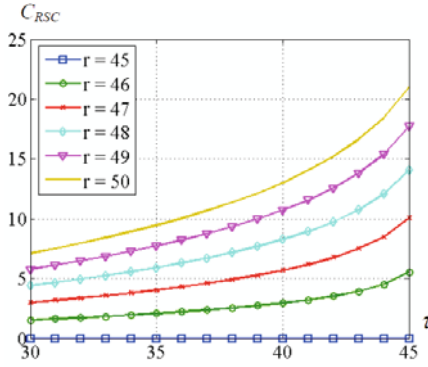


Fig. 3. Dependence of complexity of RSC decoding on τ and r

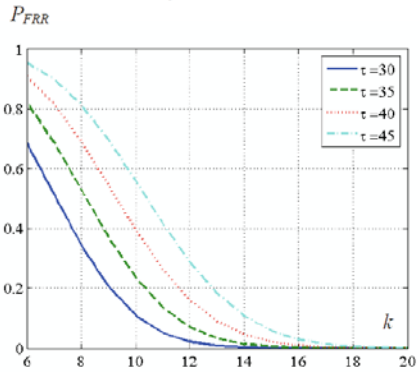


Fig. 4. Dependence of FRR on k and τ

can be used for service authentication on behalf of the user on all (permitted to this user) knots of VO.

References

1. Varetsky, Y., Rusyn, B., Ignatovych, A.: Biometric data embedding in x.509 certificates for grid systems. Informatyka w dobie XXI wieku, Technologie informatyczne w nauce, technice i edukacji pod red. A Jastriebowa 145–148 (2009)
2. Juels, A., Sudan, M.: A fuzzy vault scheme. In: Lapidoth, A., Teletar, E. (eds.) Proc. IEEE Int. Symp. Information Theory, vol. 408 (2002)

Human Vein Pattern Segmentation from Low Quality Images – A Comparison of Methods

Rafał Kabaciński and Mateusz Kowalski

Poznan University of Technology,
Intitute of Control and Information Engineering
ul. Piotrowo 3a, 60-965 Poznań
e-mail: rafal.kabacinski@gmail.com, cctp@op.pl

Summary. In this paper we propose two methods of human vein pattern segmentation from low quality images, called frequency high pass filtration and local minima analysis, witch are compared with the often used local thresholding algorithm. These methods are evaluated using manually assigned bifurcation points which is also proposed in this paper. Evaluation was carried out on 480 collected images, and shows that proposed methods are worth to consider in human vein pattern segmentation.

1 Introduction

Biometrics is the science testing methods for people identification on the basis of their physical or behavioral features [1]. Suggestions given by dr K. Shumizu from Hokkaido University in 1992 that a vein pattern can be used as a biometric feature gain increasing interest form human authentication researchers. Among other human characteristics, the blood vessels pattern has many advantages. It is different for every human, does not change with age, can be acquired contactless, and requires blood in veins, to be registered, which solves the liveness test problem.

Hemoglobin absorbs near infrared (NIR) radiation, while human skin passes it through. As an effect by using NIR sensor, image showing dark subcutaneous vascular net on a bright background of other tissue is obtained. Such a solution was introduced in 1930' in dermatology and angiology studies. Since popular CCD and CMOS sensors cover NIR part of the electromagnetic radiation spectrum, vein patterns can be captured easily.

The Important step in human vein pattern recognition is to properly extract a vein pattern form the background. In this paper two methods, alternative to often used local thresholding algorithm are proposed. To compare them, a dataset of NIR hand pictures was collected and evaluation method is proposed. Results demonstrate that these two algorithms are worth of considering in hand vein pattern segmentation and further research. This paper

is organized as follows: Section 2 describes the database and the system used for collecting it. Section 3 presents the first of the proposed methods, based on Discrete Fourier Transform. The second gradient based algorithm is introduced in Section 4. Section 5 shows the evaluation method, and provides the results of evaluation. Section 6 contains some concluding remarks and proposals for further research.

2 Dataset Collection

In order to collect data needed for verification of methods described further in this paper, an appropriate NIR Image Acquisition System (NIAS) was designed. Assumptions were: low cost, modularity, quick reconfigurability and component exchange, compactness and mobility.

The NIAS (Fig. 1) is composed of the following parts: the image acquisition module with a low-cost USB web camera, the lightning section consisting of IR LED lamps placed around the camera lens and directed backward to the object. Paper blends are added to preform diffusion-like reflection of NIR radiation. Uniform lightning generated in this way reduces light spot effects rising segmentation difficulties. For a full control of lightning conditions both lightning and image acquisition modules are housed in a case-box with a window facing the camera lens. During the acquisition process a volunteer places his/her palm or wrist above the window thus cutting off the external radiation disturbances. IR lamps are grouped into several sections with different radiation wavelengths: 780, 850, and 880 nm. Thus it is possible to test a wavelength influence on the image quality and segmentation results. After the preliminary test it was decided to use 850 nm wavelengths, although the test did not exposed significant differences between results obtained with other wavelengths. There are two regions of interest (ROI) registered: the dorsal aspect of a palm and the bottom of a wrist of left and right hands. For both registration modes quasi-rigid positioning systems were designed to stabilize each subject's palm or wrist over the window. Positioners are called quasi-rigid as they do not impose the exact hand position – there are no elements holding or blocking the palm or the wrist.

The process of dataset collection was taken on a group of 95 volunteers 22-25 years old with strong majority of males. 10 sets with at least 12 taken pictures per ROI (480 images) were selected. Pictures (in resolution 768×1024) were taken in three or four series with a interval not shorter then one week to examine a short time variation of the image acquisition results.

3 The Method Based on Discrete Fourier Transform

Discrete Fourier Transformation (DFT) is the fundamental tool in signal processing but it is relatively seldom used in image processing because of difficulties with image interpretation in the frequency domain [3]. DFT in image

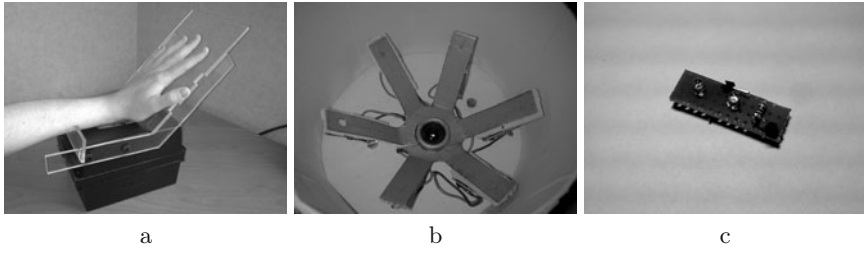


Fig. 1. a- device with the wrist positioning system and a properly positioned hand, b- lightning section, c- single IR LED lamp

processing is frequently used for the linear filter design. The first method proposed in this paper involves DFT and filtration in frequency domain. Before we move to the description of the method we recall some elements of DFT theory. It is assumed that image is an array of complex numbers where real part is an integer value within a range $< 0, 255 >$ representing pixel value and imaginary part is equal to zero.

For an array (image) \mathbf{X} of size $M \times N$ DFT is defined by the formula:

$$f(k, l) = \sum_{m=0}^{M-1} \left(\sum_{n=0}^{N-1} x(m, n) e^{-j \frac{2\pi}{N} nl} \right) e^{-j \frac{2\pi}{M} mk} \quad (1)$$

and an inverse DFT (iDFT) is given by the formula:

$$x(m, n) = \frac{1}{N} \sum_{l=0}^{N-1} \left(\sum_{k=0}^{M-1} f(k, l) e^{j \frac{2\pi}{M} mk} \right) e^{j \frac{2\pi}{N} nl} \quad (2)$$

where: $x(m, n)$ is an element of array \mathbf{X} , $f(k, l)$ is an element of array \mathbf{F} (image spectrum), $j = \sqrt{-1}$, m, n are indices of an image sample and k, l are indexes of corresponding spatial frequencies ($0 \leq m, k \leq M - 1$; $0 \leq n, l \leq N - 1$).

In textbooks about frequency filter design it is emphasized to position the filter transfer function symmetrically to the DC component of the spectrum. After DFT a spectrum is not symmetrical as a DC component appear in the top-left corner so shifting is to be applied to the spectrum or to the filter before multiplication.

The Gaussian filter performing high-pass filtration in frequency domain here in this paper is an array covering the image spectrum and filled with samples within the range $< 0, 1 >$ computed by the formula:

$$h(n_1, n_2) = 1 - e^{-\frac{(n_1^2 + n_2^2)}{2\sigma^2}} \quad (3)$$

where n_1, n_2 are pixel coordinates in local shame.

The image spectrum in the frequency domain is complex thus the 2D filter transfer function should have the following form:

$$H(k, l) = h(k - x_{dc}, l - y_{dc}), \quad (4)$$

where x_{dc}, y_{dc} are spatial coordinates of a DC component and h is given by (3).

The vein pattern is a structured element on the image so it should have its visible representation in frequency components. So the goal is to find right frequencies and suppress them in order to make vessels darker w. r. t. their surrounding. Segmentation process applied can be described with the following steps:

1. For the raw image presented in Fig. 2a the spectrum is calculated according to (1),
2. Spectrum is multiplied by Gaussian high-pass filter given by (4) with $\sigma = 0.8$ chosen experimentally (Fig. 2b),
3. New image (Fig. 2c) is computed with (2),
4. Raw result of iDFT is smoothed with 9×9 block filter (see Fig. 2d,e),
5. At the end, the image is globally thresholded with threshold equal to 0 (without greyscale rescaling).

Parameters in this method are: σ for the Gaussian function and size of the block filter. The result can be seen in Fig. 2f.

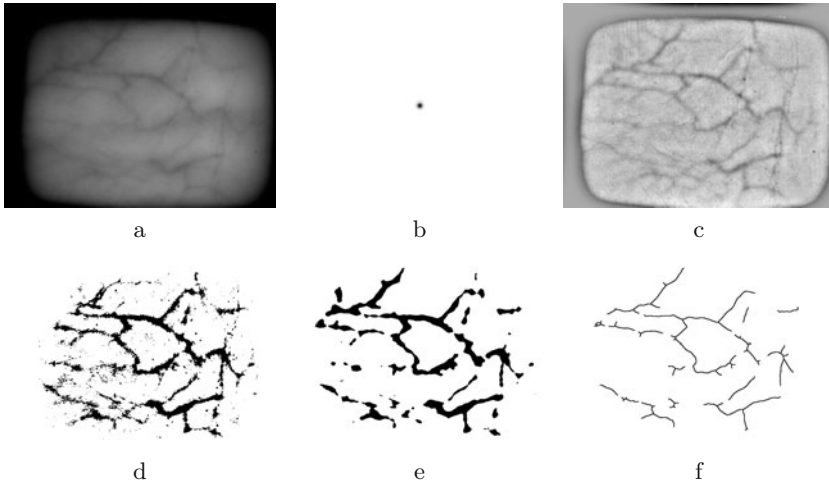


Fig. 2. Top row shows the frequency filtration process: a - input image, b - filter for filtration in frequency domain, c - filtration result in greyscale; bottom row shows postprocessing stages in inverted colors: d - frequency filtration result (without scaling to greyscale) global thresholded at zero, e - image d with average filtration for smoothing applied before thresholding, f - thinning results.

4 The Gradient-Based Segmentation Method

Blood vessels in an NIR image are darker than other tissues, therefore their direct extraction can be based on finding local minima of the image function. To reduce a calculation cost, local minima can be searched for in rows and columns only. An exemplary image and the profile of its middle row are presented in Fig. 3. To remove noise, a 2D low-pass Gaussian filter is used in the loop. Stop conditions for this loop are: the elimination of every single-pixel-wide minimum in three predefined rows and columns (in $1/4$, $1/2$, and $3/4$ of the image width and height), or trespassing of a predefined number of iterations. An exemplary row, after filtering is shown in Fig. 3. Local

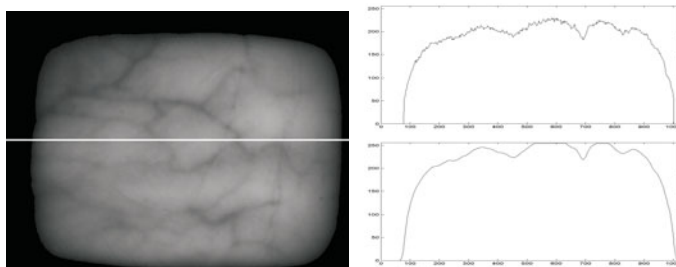


Fig. 3. Oryginal image, with marked row (left). Row beafore filtration (right top). Row after filtration, and scaling (right bottom).

minimum and maximum detection is based on finding zero-crossing points of a directional derivative of the image function. This method does not involve large memory, because only a value of preceding pixel is needed to determine the pixel gradient. Thus no additional temporary image is necessary. A result of local minimum detection is the set of separated points presented in Fig. 4. Many of these points result from the noise, not completely removed by filtering, but those points representing blood vessels form curvilinear clusters. Using the morphological closing w. r. t. this image would cause many parasitic branches after thinning the blobs. Therefore, after locating a local minimum, 10 pixels before and after this minimum are set to 1. The result image can be seen in Fig. 4. Morphological erosion is then used to erase lines isolated from the non-vein minima. In the next step morphological closing is performed on the image, then second erosion is applied to improve the continuity and smoothness of skeleton, and to remove smaller artifacts. Finally the thinning algorithm described in [4] is used. The resulting image is presented in Fig. 4. This method requires three parameters for filtration process (kernel size, σ , and the maximal number of iterations), one for the method itself (the length of minimum marking lines), and one for postprocessing (the size of a morphological closing). In total 5 parameters are needed in the method.

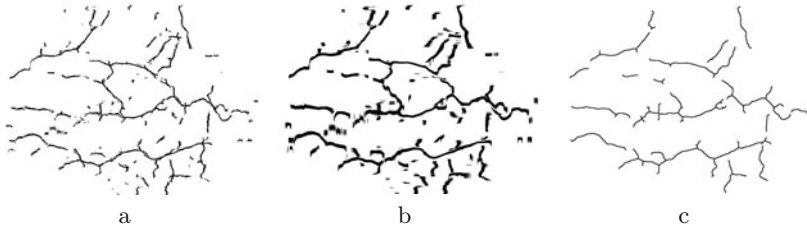


Fig. 4. Gradient results in inverted colors: a - Local minimums marked as dots, b - and marked by lines, and c - result of gradient algorithm

5 Results of the Experiment

The main step in human vein pattern recognition is to properly extract a vein pattern from its background. Problem is even more difficult with low quality and high resolution images like those discussed in this paper. Some simple methods based on histogram analysis like global thresholding do not provide satisfying results. Therefore, researchers often decide to use an adaptive threshold (local thresholding – see: [2, 5]). This algorithm needs filtration in preprocessing stage. In this case median, and Gauss filters are used, which means that 3 parameters are to be set. Next 2 parameters are used in the algorithm. To improve the output, the morphological erosion is applied, whose mask size is the last parameter. In total 6 parameters are needed to be chosen.

To evaluate algorithms described in previous Sections [3] and [4], we compare them with this local thresholding method. Evaluation is based on comparing bifurcation points assigned manually (ground truth) with those obtained by compared methods. The dataset consisting of hand images of ten subjects, with twelve pictures per one ROI (totally 480 images), was used. Bifurcation points were assumed to be properly extracted if they lay within five-pixel or ten-pixel distance apart from each other. Performance indexes calculated for each picture were: sensitivity S (see [6]) showing a percent of points correctly recognized by the method and a positive predictive value percent PPV (see [7]) showing how many points marked by a method is marked properly. They were calculated by the following formulas:

$$S = \frac{TP}{TP + FN} * 100\% \quad (5) \quad PPV = \frac{TP}{FP + TP} * 100\% \quad (6)$$

where: FN (false negative) is the number of points marked manually and not marked by a method, FP (false positive) is the number of points not marked manually and marked by a method, TP (true positive) is the number of points marked manually and marked by a method.

Statistical parameters characterizing a level of consistency between points extracted with each method and points marked manually were calculated for the entire population of images. The results are gathered in Tables [1] – [3].

Table 1. Comparison of methods: sensitivity percent according to (5)

Method	Hand				Wrist			
	minimum	maximum	average	median	minimum	maximum	average	median
radius = 5								
LT	0	55	21.91	19.23	0	63.64	17.68	14.29
Fourier	0	50	26.76	25	0	57.14	19.18	15.38
Gradient	0	47.37	21.49	19.05	0	63.64	18.41	15
radius = 10								
LT	24	92.31	49.40	46.67	0	85.71	42.18	38.46
Fourier	13.64	77.42	49.93	46.88	0	87.50	41.11	36.84
Gradient	18.52	69.57	43.10	40	0	85.71	40.29	37.50

Table 2. Comparison of methods: positive predictive value percent according to (6)

Method	Hand				Wrist			
	minimum	maximum	average	median	minimum	maximum	average	median
radius = 5								
LT	0	16.92	6.43	5.48	0	17.24	4.75	3.51
Fourier	0	62.50	17.48	13.16	0	45.45	12.21	8.33
Gradient	0	27.78	9.06	7.25	0	23.08	4.93	3.70
radius = 10								
LT	4.92	29.09	14.53	12.75	0	27.91	11.13	9.43
Fourier	5.96	87.5	32.16	27.27	0	75	26.06	20
Gradient	3.57	43.33	17.95	15.79	0	31.25	10.84	8.96

Table 3. Comparison of methods: number of true positive (TP)

Method	Hand				Wrist			
	minimum	maximum	average	median	minimum	maximum	average	median
radius = 5								
LT	0	12	4.92	4	0	7	2.19	2
Fourier	0	14	6.12	5	0	8	2.30	2
Gradient	0	12	4.89	4	0	10	2.25	2
radius = 10								
LT	3	21	11.17	11	0	13	5.11	5
Fourier	2	24	11.41	11	0	13	4.95	4
Gradient	3	20	9.85	9	0	13	4.86	4

In these Tables methods are denoted as follows: LT - the local thresholding, Fourier - the DFT-based method, Gradient - the Gradient-based method.

Number of TP points found by all methods are comparable, thus also sensitivity percent's are similar. Although LT algorithm produces many FP points, which increase possibility that certain image will be identified wrongly. Therefore in table 2 DFT based Method outperforms LT, scoring almost three

times better results. Also Gradient based algorithm in many cases leads to better results than LT. Fourier needs also the smallest number of formal parameters, but its calculation cost is the highest. Gradient algorithm is fast, but needs almost the same number of parameters as LT.

6 Conclusions

In this paper two methods of vein pattern extraction from a high-resolution, low quality NIR image were introduced and compared with often used LT algorithm. Experiments were run on the data-set acquired with the system based on a standard USB camera. The evaluation method used manually assigned bifurcation points as ground truth. The experiment demonstrates that the proposed methods give comparable sensitivity results, while giving less false positive points than the local thresholding based segmentation algorithm. They also require less formal parameters. In case of the DFT based method, only a high calculation cost is inferior to LT. Future research should be focused on reducing the number of parameters required by methods and improving their performance. It is also important to compare their efficiency in person identification.

References

1. Ratha, N.K., Senior, A., Bolle, R.M.: Tutorial on automated biometrics. In: Proc. Int. Conf. Advances in Pattern Recognition, Rio de Janeiro, Brazil (2001)
2. Wang, L., Leedham, G., Cho, S.-Y.: Infrared imaging of hand vein patterns for biometric purposes. *IET Comput. Vis.* 1(3-4) (2007)
3. Tadeusiewicz, R., Korohoda, P.: *Komputerowa analiza i przetwarzanie sygnałów*. Wydawnictwo Fundacji Postępu Telekomunikacji, Kraków (1997)
4. Lam, L., Lee, S.-W., Suen, C.Y.: Thinning Methodologies - A Comprehensive Survey. *IEEE Transactions On Pattern Analysis And Machine Intelligence* 14(9), 879 (1992)
5. Shahin, M., Badawi, A., Kamel, M.: Biometric Authentication Using Fast Correlation of Near Infrared Hand Vein Patterns. *International Journal of Biological and Life Sciences* (2006)
6. Lopes, H.S., Coutinho, M.S., Lima, W.C.: An Evolutionary Approach to Simulate Cognitive Feedback Learning in Medical Domain. In: *Genetic Algorithms and Fuzzy Logic Systems*, pp. 193–207. World Scientific, Singapore (1997)
7. Altman, D.G., Bland, J.M.: Diagnostic tests 2: Predictive values. *British Medical Journal* 309(6947), 102 PMID 8038641 (1994)

A Modified Algorithm for User Identification by His Typing on the Keyboard

Piotr Panasiuk and Khalid Saeed

AGH University of Science and Technology
e-mail: panasiuk@agh.edu.pl, saeed@agh.edu.pl

Summary. In this paper the authors modify their previous kNN algorithm and present a modification to improve the algorithm by considering key inner and interclass distinguishability. The suggested approach is tested on a large group of individuals with data gathered over Internet using browser-based WWW application. The obtained results are promising and encouraging for further development in this area.

1 Introduction

With the spread of the Internet access and increasing computational power of the personal computers, there comes expanding risk of our private data to be stolen. Not only highly restricted access systems are in danger, but also the personal desktop computers are. With more computational power comes the ease of brute-force hacking and the expanding risk of stealing our private data. In various security systems there are three authentication methods: memory-based, token-based and biometrics-based authentication. Memory-based authentication is a valid user's knowledge-based method. Examples are password, PIN number or where to go, what to say. The second authentication method token-based is to verify the identity on the basis of something that the user has, e.g. a key, a chip/magnetic card that confirms the user's identity. Sometimes, those methods can be combined as the case of payment cards. Another class of security methods are biometrics-based. These methods use physical and behavioral human features to confirm user's identity. The algorithm presented in this paper belongs to this group of security solutions.

Securing data with passwords appears to be insufficient, especially when there are lots of users who think this problem does not concern them and they use easy passwords like 123456, the user's name, etc. As a typical secure password we consider a phrase with at least six characters length (this value grows with time) and consists of small and capital letters, digits and symbols combined altogether and does not appear in a dictionary. Usually it is hard to remember such difficult words. As a consequence they are easy to forget,

or as in the previous situation, passwords are written down somewhere near the workstation. This last solution is very risky in practice, but quite popular in some offices. Therefore, there still exists the need for a simple password resistant to brute-force method. Combining memory-based and biometrics-based authentication methods a keystroke dynamics verification approach has been developed. It is convenient for the user because the password may be simple but hard for the impostor to break. It is also a noninvasive method as the user randomly types his default password. Besides, knowing what to type, the potential attacker has to know how to type it and this is not easy. What is more, keystroke dynamics can make brute-force hacking methods impossible or much slower. Even if someone emulates the user's average typing speed, the brute-force will not be effective enough to finish the task if it accidentally matches both the phrase and typing pattern, which is assumed to be possible.

2 Other Works on Keystroke Dynamics

Keystroke dynamics is quite a new biometric technique, but there have already been made some trials to utilize the keyboard using to identify people. Early researches in this technique began in late 1970's. The first work that intentionally examined the uniqueness of typing style was driven in IBM laboratories in 1975 [1]. This encouraged many researchers to examine this field. In 1979 SRI International (formerly Stanford Research Institute) developed the first hardware-based implementation of keystroke dynamics algorithm [2]. In 1980 under foundation of National Bureau of Standards, SRI achieved about 98% accuracy of their approach [3]. This research officially confirmed that typing pattern is unique for one person and could be distinctive.

In 1980 a similar research [4] was made by RAND Corp. The experiment was conducted on population of seven secretaries. The obtained results had errors of $FRR=5.5\%$ and $FAR=55\%$. Also worth to mention are such works as [5] and [6] made by Leggett et al., which were the first attempts to continuous verification. Hussien et al. [7] were the first who used neural networks in keystroke dynamics. This started widely usage of neural networks in keystroke dynamics research. All of those methods return good results. Some researchers obtained errors rates less than 1%. Obaidat and Sadoun wrote an article [8] in which they compared a number of neural network algorithms. Their best result for 15 users had 0% FAR and FRR error rates.

A very interesting approach was suggested by Bergadano et al. [9]. As examined features they took trigrams (time between pressing the first and the third key). They calculated trigrams from the examined word and stored it as a vector. In the next step the test vector was compared with the training vectors and the "degree of disorder" was calculated. For "degree of disorder" the authors considered the amount of differences in both vectors. Times were not compared between two vectors, only sequence similarity was considered. The best accuracy of their approach for 44 users was 0,13% FAR and 0% FRR.

In search of new ways of keystroke analysis and new features of the resulted signals there has been designed and built a touch-sensitive keyboard [10]. This research shows that pressure force used during a text typing is a very helpful feature that can improve the algorithm accuracy. However, this kind of device is hard to get and the cost of its production will be much more expensive than a simple keyboard. Maybe when pocket and tablet PCs with virtual keyboards become more popular, this feature will be used more widely in the field of keystroke dynamics.

Another interesting research based on pressure force was conducted in 2009 [11]. Scientists had changed a simple laptop touchpad into a touch-sensitive numerical keyboard. This device was like ATM's keyboard. However, instead of typing some text, users had to type their phone number. There was 10 users typing 10-digit numbers. What is interesting, EER (Equal Error Rate) for only flight times was 40%, for dwell times was 35% and when only pressure was considered ERR was equal 1%. These results are mainly affected by the short "phrase" and the only 10 characters involved (from 0 to 9). What is more, the numbers were different for every user, because they were their phone numbers. Examining free text for such little information could not give good classification accuracy, though the results show great advantage of using pressure force analysis in keystroke dynamics.

3 Authors' Suggested Approach

This work is the continuation of the authors' previous approaches to keystroke dynamics. The first approach [12] was aimed at free-text identification applications. However, when fixed text was used [13], the results were not so good. In this paper we present a modification which improves the classification accuracy of the fixed text approach.

Data have been gathered in non-supervised way with the web-based platform located in [14]. Authors' intention was to create a web-based data gathering system using the most popular technologies (html, JavaScript) independent of the operating system so that everyone with Internet access, anywhere in the world can contribute to the research. This platform exists in two language versions: Polish and English. However, English version, unfortunately, showed less interest, so in this approach we mainly used the Polish version samples only.

Samples consist of five phrases that everyone has their unique features. In both language versions are the same dependencies in sample selection. Each phrase in a sample is stored in the database as a series of key events written as a text.

At the beginning we read the SQL file and load it into the testing subprogram. Each keyboard event in the database is recalculated from the time a key is pressed or released to the flight time and dwell time. Flight time is the time between releasing a key and pressing the second. Dwell time is the time when a specific key is in pressed state.

After loading the database file into the testing platform backspace and delete keys are removed from samples with affected keys information. Later are removed samples with different events count than the most common (web browser/JavaScript implementation disturbance). After this, the users with less number of samples than the training set size plus at least one test sample per user are removed from the database. The next step is splitting the remained database into test and training sets. Previous steps provided constant count of training samples per user. In the following step every test sample in the remained database is being classified. For this purpose we calculate the similarity distance between test sample and every training sample in the training data set. The distance is calculated using Manhattan metrics. It is the sum of differences between flight and dwell times of the corresponding keystrokes as in (1) to (3).

$$fDist(a, b) = \sum_{i=1}^{n-1} |flight_{ai} - flight_{bi}| \quad (1)$$

$$dDist(a, b) = \sum_{i=1}^n |dwell_{ai} - dwell_{bi}| \quad (2)$$

$$dist(a, b) = fDist(a, b)p - dDist(a, b)(1 - p), \quad (3)$$

where $fDist$ is the distance of *flight* times and $dDist$ is the distance of *dwell* times, a and b are the compared test and training samples, and i is the index of the corresponding pressed key. There is n dwell times and $n-1$ flight times. Additionally, we use p value. It is the ratio of flight times against dwell times. As we have list of distances we use the k nearest neighbour classifier with voting mechanism. The shortest distance gains a weight equal to k , the next in $k-1$ and so on until the weight is equal to 1. Those weights have assigned user id and are considered as weighted votes on particular user. Finally, we add the values achieved by the specific user, and the winner of the algorithm is the user who gained the biggest weight.

This approach has been improved by using weights for each flight and dwell time in a phrase. This was possible because the system was operating on fixed-text phrases and it was known which keys are in use and what times are available. Time weight was calculated using the training samples. The first step to calculate the weights was calculating the mean values and variances for each time in every specific user's sample. After that for each key, the mean set separations were calculated for user i and j using Fisher's discriminant (4).

$$F_{ij} = \frac{|mean_i - mean_j|}{stdDev_i + stdDev_j} \quad (4)$$

$$W = \frac{\sum_{i=1}^n \sum_{j=1; j < > i}^n F_{ij}}{n(n-1)} \quad (5)$$

The last step was to calculate the mean value from all separation factors for each key (5), and after normalization those results were considered as weights for each duration time.

4 Experimental Results

In the conducted experiments only user id and user samples were used. No other database feature was taken into account. In Fig. 1 one can see that classification accuracy of the improved algorithm is higher than the previous approach [13]. The tested phrase was "Litwo! Ojczyzno moja! Ty jesteś jak zdrowie." which consists of 44 characters. The presented results are obtained for the best p value. We do not have the p value setting automated yet. K was set to 2 in this experiment and the training samples count was the same for all users and set to the values from 1 to 9. One can see a big loss of the new algorithm accuracy for 1 training set. This is because the new algorithm uses standard deviation; it is zero here and causes a big error. For 2 training samples the results are much better. The new algorithm is winning with the simple kNN algorithm. This proves that considering inter and intra class distances improves the algorithm accuracy. The difference is especially noticeable when we have small number of training data items. It may be caused by the weights that get more equal to each other in the case of calculating more training samples. With such many samples, some distinguishable user features may be blurred with noise.

The users taken into experiment were randomly selected from the database and each result presented below is the mean value of at least 10 executions per experiment at specific settings.

After conducting an experiment for different k nearest neighbours value it revealed that the best value for our current algorithm is 2. For $k=1$ we have noticed significant accuracy loss and for values above 3 the accuracy was decreasing. Comparing with the previous algorithm, the results of classification show a higher accuracy.

As can be seen in Fig. 2 the algorithm stability for different flight time weight values is improved. This despite the fact that when we consider only flight times or dwell times (weight is equal to 0 or 1) there is noticeable accuracy loss. In the future this solution will reduce classification error as a side effect of the automatically tuning flight time weights.

Another experiment was conducted to examine the dependance of accuracy on training samples used for a phrase that is hard to pronounce. In English this type of phrases is called "tongue twisters". This phrase ("Stół z powyłamywanymi nogami.") consists of 29 characters and as one can see it consists of small, capital and diacritic letters. Results are similar to longer



Fig. 1. Results of algorithm classification accuracy for 44-character phrase, 16 users, $k=2$ and various training samples used in the learning process

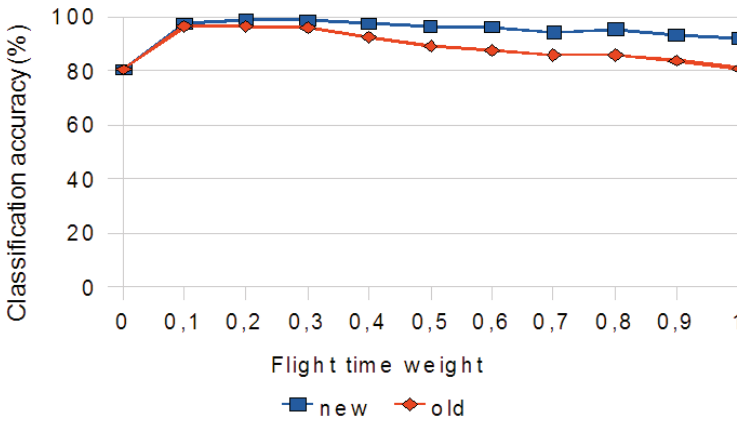


Fig. 2. Classification accuracy for different flight time weight values. The test was conducted on 44-character phrase with 16 users in database, 9 training samples and $k=2$.

phrase results. The best classification accuracy gained for this phrase was 91.4% for the new method and 89.8% for the previous one with 16 users in the database and 9 training samples used.

In Fig. 3 one can see the relation between the user count and the identification accuracy. Maximum number of users is 79 because many users had at least 4 correct samples on their accounts. Three of them were used for training, and at least one was used during the test.

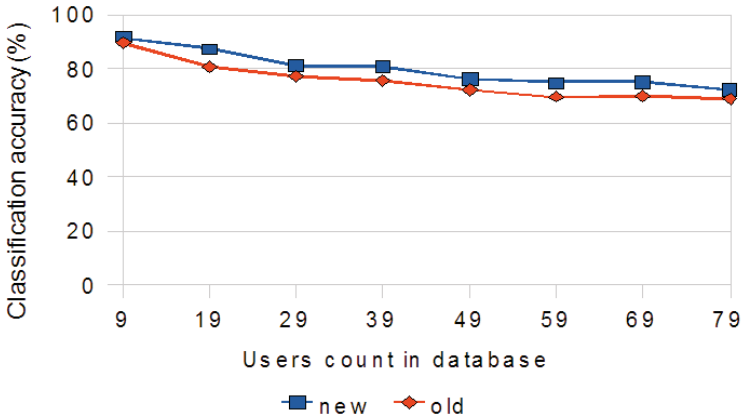


Fig. 3. Classification accuracy for different class amount. Test was conducted with 44-character phrase, 3 training samples and $k=2$.

5 Conclusions and Future Work

In this paper we presented an algorithm to identify the computer user with keystroke dynamics. Data was gathered via Internet. The system was available for all users and non-supervised. Samples stored in our database were fixed, so each user had to input the same phrases. After cleaning the samples from errors and web browser artefacts we started the tests. For 9 training samples, $k=2$ and 16 users taken into account, k NN classification method with simple voting mechanism returned good identification accuracy at the level of 96.44%. The simple algorithm was improved by using Fisher's discriminant, which increased the classification accuracy to 98.78% keeping short operation time. Of course with more users taken into consideration, the accuracy would decrease. This is obvious because with more classes the decisions are harder to make. Tests made for 3 training samples per user allowed us to check what the identification accuracy would be for 79 classes. The success rate for this experiment was about 72.3%. This result would be much better if we had 9 training samples per each user like the experiment mentioned earlier. In comparison, randomly user guessing would get 1.27% success rate. This algorithm still needs some development, especially the flight time weight automation process, although experimental results gained for our database are very optimistic, but still higher accuracy is required.

Our future work will be focused on improving the classification method. We also need to implement the verifying algorithm constantly. This would authenticate the user in the real time during his normal, everyday work. The main goal is to create a system that would protect private data without bothering the user or even notifying their activity. After detecting the impostor user, the system can ask for the password, or without noticing it just sends

email at owner's address with the current computer localization (basing on the IP address) and the images captured from the web camera.

The authors are working on determining the user's age on the basis of keystroke dynamics.

Acknowledgement

The research is supported within SOCIONICAL - FP7 project No. 231288.

References

1. Spillane, R.J.: Keyboard Apparatus for Personal Identification. IBM Technical Disclosure Bulletin 17(3346) (1975)
2. Vacca, J.R.: Biometric Technologies and Verification Systems, p. 182. Elsevier, USA (2007)
3. Coppentrath, L.F., & Associates: 5 (2001), <http://lfca.net/Reference20Documents/Biometric20Technology200verview.pdf> (State from June 15, 2010)
4. Gaines, R., Lisowski, W., Press, S., Shapiro, N.: Authentication by Keystroke Timing: Some Preliminary Results. Rand Rep. R-2560-NSF, Rand Corp. (1980)
5. Leggett, J., Williams, G., Umphress, D.: Verifying Identity via Keystroke Characteristics. *Int. J. Man-Mach. Stud.* 28(1), 67–76 (1988)
6. Leggett, J., Williams, G.: Verification of User Identity via Keystroke Characteristics. *Human Factors in Management Information Systems*, 89
7. Hussien, B., McLaren, R., Bleha, S.: An application of fuzzy algorithms in a computer access security system. *Pattern Recog. Lett.* 9, 39–43 (1989)
8. Obaidat, M., Sadoun, B.: Verification of computer users using keystroke dynamics. *IEEE Trans. on Systems Man and Cybernetics, Part B* 27(2), 261–269 (1997)
9. Bergadano, F., Gunetti, D., Picardi, C.: User authentication through keystroke dynamics. *ACM Transactions on Information and System Security* 5(4), 367–397 (2004)
10. Loy, C.C., Lai, W.K., Lim, C.P.: Keystroke Patterns Classification using the ARTMAP-FD Neural Network. *Intelligent Information Hiding and Multimedia Signal Processing* 1, 61–64 (2007)
11. Saevanee, H., Bhattarakosol, P.: Authenticating user using keystroke dynamics and finger pressure. In: *Proceedings of the 6th IEEE Conference on Consumer Communications and Networking Conference*, pp. 1078–1079 (2009)
12. Rybnik, M., Tabedzki, M., Saeed, K.: A Keystroke Dynamics Based System for User Identification. In: *Computer Information Systems and Industrial Management Applications - CISIM*, pp. 225–230. IEEE, Los Alamitos (2008)
13. Rybnik, M., Panasiuk, P., Saeed, K.: User Authentication with Keystroke Dynamics using Fixed Text. In: *2009 International Conference on Biometrics and Kansei Engineering*, pp. 70–75. IEEE CS, Los Alamitos (2009)
14. (July 20, 2010), <http://www.kds.miszu.pl>

Multimodal Biometric Personal Authentication Integrating Iris and Retina Images

Ryszard S. Choraś

Department of Telecommunications and Electrical Engineering
University of Technology & Life Sciences
85-796 Bydgoszcz, S. Kaliskiego 7, Poland
e-mail: choras@utp.edu.pl

Summary. In this paper Iris and Retina features are combined for recognition in biometric system. In this multimodal biometric system two biometrics can be taken from the same acquisition process and image. Gabor transform to extract the features from Iris and Retina is used. Feature fusion is performed.

1 Introduction

Personal identification is crucially significant in a variety of applications. Conventional person's identification systems use keys, passwords, PIN numbers and other tokens as they are easy to use however they are insecure.

Using passwords as personal identification carries the risk as the user often forgets it or it can be a subject to fraud. Users encounter problems in terms of theft, loss, and reliance on the own (user's) memory. Biometric technology that uses human body information can decrease the risk because the biometric systems which recognise users based on their physiological and behavioural characteristics eliminate memorization and eliminate misplaced tokens.

Physiological biometrics (also known as static biometrics) are based on data derived from the measurement of a part of a person's anatomy. For example, fingerprints and iris patterns, as well as facial features, hand geometry and retinal blood vessels. Behavioural biometrics are based on data derived from measurement of an action performed by a person and, distinctively, incorporate time as a metric as the measured action. For example, voice (speaker verification).

All biometric systems work in a similar fashion:

1. The user submits a sample that is an identifiable, unprocessed image or recording of the physiological or behavioural biometric via an acquisition device,
2. This image and/or biometric is processed to extract information about distinctive features.

Biometric systems have four main components [1]: sensor, feature extraction, biometric database, matching-score and decision-making modules (Fig. 1). The input subsystem consists of a special sensor needed to acquire the biometric signal. Invariant features are extracted from the signal for representation purposes in the feature extraction subsystem. During the enrolment process, a representation (called template) of the biometrics in terms of these features is stored in the system. The matching subsystem accepts query and reference templates and returns the degree of match or mismatch as a score, i.e., a similarity measure. A final decision step compares the score to a decision threshold to deem the comparison a match or non-match.

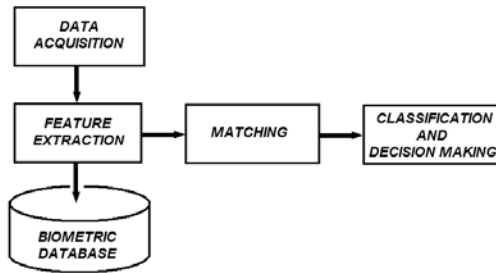


Fig. 1. Biometric system

The ideal biometric characteristics have five qualities:

1. Robust: Unchanging on an individual over time. "Robustness" is measured by the probability that a submitted sample will not match the enrolment image.
2. Distinctive: Showing great variation over the population. "Distinctiveness" is measured by the probability that a submitted sample will match the enrolment image of another user.
3. Available: The entire population should ideally have this measure in multiples. "Availability" is measured by the probability that a user will not be able to supply a readable measure to the system upon enrolment.
4. Accessible: Easy to image using electronic sensors. "Accessibility" can be quantified by the number of individuals that can be processed in a unit time, such as a minute or an hour.
5. Acceptable: People do not object to having this measurement taken on them. "Acceptability" is measured by polling the device users.

The problem of resolving the identity of a person can be categorized into two fundamentally distinct types of problems with different inherent complexities:

- (i) verification (also called authentication) refers to the problem of confirming or denying person's claimed identity (Am I who I claim to be?)

(Fig. 2). It is required that a user claims an identity in order for a biometric comparison to be performed. Once the identity is claimed, the user provides biometric data, which is then compared against his or her enrolled biometric data. To claim an identity, the user may use a username, a given name, or an ID number. The answer returned to the system is a match or not a match.

and

- (ii) identification (Who am I?) refers to the problem of establishing a subject's identity. They do not require that a user claims an identity before biometric comparisons take place. The user provides biometric data, which is compared to data from a number of users to find a match. The answer returned to the system is an identity.

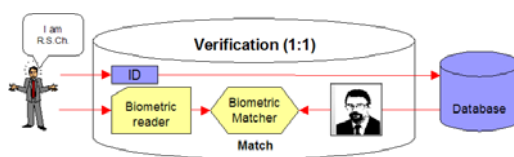


Fig. 2. Verification process

Unimodal biometric systems make use of a single biometric trait for user recognition. It is difficult to achieve very high recognition rates. A multimodal biometric system utilizes a number of different biometric templates like face, fingerprint, hand-geometry, and iris. Such system can achieve higher recognition accuracy than unimodal systems. However, a multimodal system will require longer verification time.

In this paper recognition methods are presented for authentication of a person on the basis of a feature vector derived from an iris and retina input images.

2 Iris and Retina Recognition

The iris and retina features are combined for recognition in biometric system. In this multimodal biometric system two biometrics can be taken from the same acquisition process. Invariant features are extracted from the signal for representation purposes in the feature extraction subsystem. During the enrolment process, representation (called template) of the biometrics in terms of these features is stored in the system. Feature fusion is performed. The matching subsystem accepts query and reference templates and returns the degree of match or mismatch as a score, i.e., a similarity measure. A final decision step compares the score to a decision threshold to deem the comparison a match or non-match.

Iris texture patterns are believed to be different for each person, and even for the two eyes of the same person. It is also claimed that for a given person,

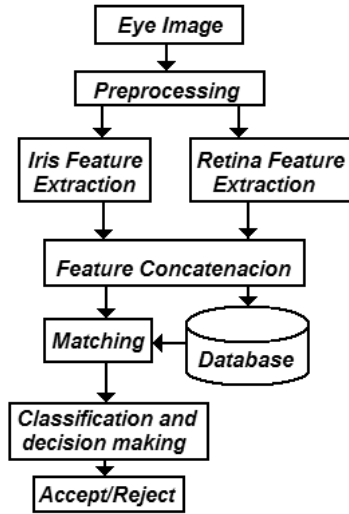


Fig. 3. The block diagram of the multimodal biometric system based on fusion of iris and retina features

the iris patterns change little after youth. The iris is the coloured portion of the eye that surrounds the pupil. Its combination of pits, striations, filaments, rings, dark spots and freckles are the very accurate means of biometric identification [6]. Its uniqueness is such that even the left and right eye of the same individual is very different.

Retina recognition technology captures and analyzes the patterns of blood vessels on the thin nerve on the back of the eyeball that processes light entering through the pupil. The retina biometric analyzes the layer of blood vessels located at the back of the eye. The blood vessels at the back of the eye have a unique pattern, from An eye to an eye and a person to a person. The retina, a layer of blood vessels located at the back of the eye, forms an identity card for the individual under investigation.

A major approach for iris and retina recognition today is to generate feature vectors corresponding to individual iris and retina images and to perform these matching based on some distance metrics [6, [11].

Random iris and retina patterns can be seen as texture, so many well-developed texture analysis methods can be adapted to recognize these images. Gabor filters are used to extract the iris and retina features. Additionally we used Haralick texture features for the iris template and some geometrical features for retinal templates.

The general functionality of the 2D Gabor filter family can be represented as a Gaussian function modulated by a complex sinusoidal signal [12] (Fig. 4).

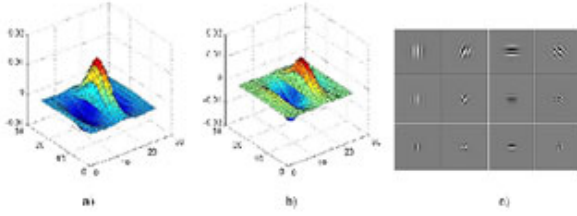


Fig. 4. Real (a) and imaginary (b) parts of Gabor wavelets and Gabor kernels with different orientations (c)

In our work we use a bank of filters built from these Gabor functions for texture feature extraction.

The two-dimensional Gabor filter is defined as

$$Gab(x, y, W, \theta, \sigma_x, \sigma_y) == \frac{1}{2\pi\sigma_x\sigma_y} e^{\left[-\frac{1}{2}\left(\left(\frac{x}{\sigma_x}\right)^2 + \left(\frac{y}{\sigma_y}\right)^2\right) + jW(x \cos \theta + y \sin \theta)\right]} \quad (1)$$

where $j = \sqrt{-1}$ and σ_x and σ_y are the scaling parameters of the filter, W is the radial frequency of the sinusoid and $\theta \in [0, \pi]$ specifies the orientation of the Gabor filters.

2.1 Iris Recognition

The initial stage deals with iris segmentation. This consists of Localized iris inner (pupillary) and outer (scleric) borders. Robust representations for iris recognition must be invariant to changes in the size, position and orientation of the patterns. To each pixel of the iris, a pair of real coordinates (r, θ) , where r is on the unit interval $[0, 1]$ and θ is an angle in $[0, 2\pi]$.

Because most of the irises are affected by upper and lower eyelids, the iris is divided into two rectangular (Fig. 5a) or two angular sectors (Fig. 5b) having the same size. The blocks of interest (ROI) should be isolated from the normalized iris image.

Two sectors with Fig. 5 are transformed into a normalized rectangular blocks each of size 32×128 to achieve size independent iris recognition. We uses in our iris recognition systems features based on Gabor functions analysis and texture features based on the Haralick's approach.

Gabor filtered output of the image is obtained by the convolution of the image with Gabor functions. Given an image $F(x, y)$, we filter this image with $Gab(x, y, W, \theta, \sigma_x, \sigma_y)$

$$G(x, y) == \sum_k \sum_l F(x - k, y - l) * Gab(x, y, W, \theta, \sigma_x, \sigma_y) \quad (2)$$

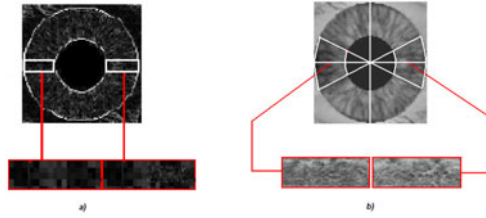


Fig. 5. The iris ROI

The magnitudes of the Gabor filters responses are represented by three moments

$$\mu(x, y) = \frac{1}{XY} \sum_{x=1}^X \sum_{y=1}^Y G(x, y) \quad (3)$$

$$std(x, y) = \sqrt{\sum_{x=1}^X \sum_{y=1}^Y (|G(x, y)| - \mu(x, y))^2} \quad (4)$$

$$Skew = \frac{1}{XY} \sum_{x=1}^X \sum_{y=1}^Y \left(\frac{G(x, y) - \mu(x, y)}{std(x, y)} \right)^3 \quad (5)$$

The feature vector is constructed using $\mu(x, y)$, $std(x, y)$ and $Skew$ as feature components.

Gabor filters worked as local bandpass filters and each filter is fully determined by choosing the four parameters $\{\theta, W, \sigma_x, \sigma_y\}$. Assuming that N filters are needed in an application, $4N$ parameters need to be optimized. The orientation parameter θ should satisfy $\theta \in [0, \pi)$. W is the radial frequency of the Gabor filter and is application dependent. σ_x and σ_y are the effective sizes of the Gaussian functions and are within the range $[\sigma_{min}, \sigma_{max}]$.

The iris features information is extracted based on the Haralick's approach [13]. The co-occurrence matrixes $P_{\delta, \theta}(x, y)$ are bi-dimensional representations showing the spatial occurrence organization of the gray levels

Table 1. Mean and standard deviation Gabor phase iris image

Iris image	$\mu(x, y)$	$std(x, y)$
$\theta = 0$	223,93	33,86
$\theta = 45$	223,84	33,75
$\theta = 90$	224,02	34,38
$\theta = 135$	225,35	32,40
$\theta = 45$ blob	215.18	27,33

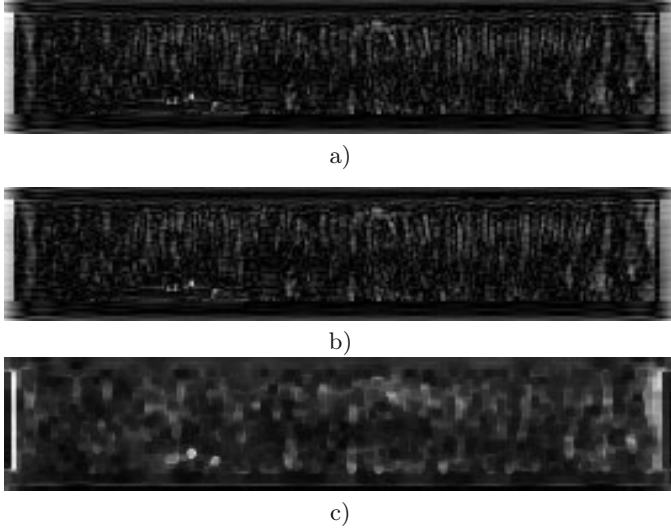


Fig. 6. Gabor transform of iris image: (a) power; (b) phase, and blob detection results for Gabor phase image (c)

in an image. They represent a bi-dimensional histogram of the gray levels, where fixed spatial relation separates couples of pixels, defining the direction and distance (δ, θ) from a referenced pixel to its neighbour.

The features are:

1. Second Angular Moment

$$SAM = \sum_{x=1}^k \sum_{y=1}^l [P_{\delta, \theta}(x, y)]^2 \quad (6)$$

2. Contrast

$$Con = \sum_{x=1}^K \sum_{y=1}^l (x - y)^2 P_{\delta, \theta}(x, y) \quad (7)$$

3. Correlation

$$Corr = \frac{\sum_{x=1}^k \sum_{y=1}^l [xy P_{\delta, \theta}(x, y)] - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (8)$$

4. Inverse Differential moment

$$IDM = \sum_{x=1}^k \sum_{y=1}^l \frac{P_{\delta, \theta}(x, y)}{1 + (x - y)^2} \quad (9)$$

5. Entropy

$$E = - \sum_{x=1}^k \sum_{y=1}^l P_{\delta,\theta}(x, y) \log P_{\delta,\theta}(x, y) \quad (10)$$

2.2 Retina Recognition

To represent retinal characteristic we using luminance component (Y) from YC_bC_r (YIQ) color space (Fig 7).

We used Gabor filters with different scales and orientation to detect blood vessels in image of the retina.

The normalized retinal image (Y components) are divided into blocks (Fig. 8). The size of each block in our application is $k \times l$ ($k = l = 20$). Each block (Fig. 9) is filtered with equation (2).

The magnitudes of the Gabor filters responses are calculated using eqs. (6) - (10).

Finally, a feature vector is constructed for each block using these components.

Table 2. Texture parameters iris image

Left ROI Iris Image		
Parameter	θ	$\delta = 1$
ASM	0 and 180	0.012
	90 and 270	0.009
Con	0 and 180	911.819
	90	1413.594
	270	2891.061
Corr	0 and 180	3.182E-4
	90	2.791E-4
	270	1.729E-4
IDM	0 and 180	0.262
	90 and 270	0.161
E	0 and 180	7.302
	90	7.791
	270	7.664

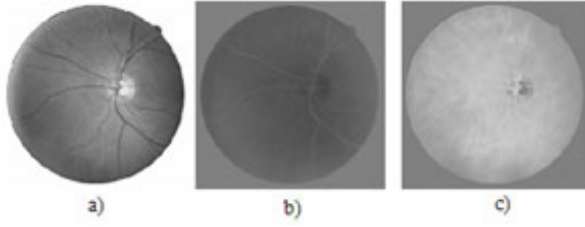


Fig. 7. Retina image in YC_bC_r color space: a) Y component and components C_b (b), C_r (c) respectively.

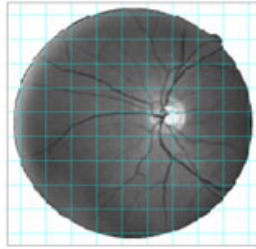


Fig. 8. Original block retinal images (Y component)

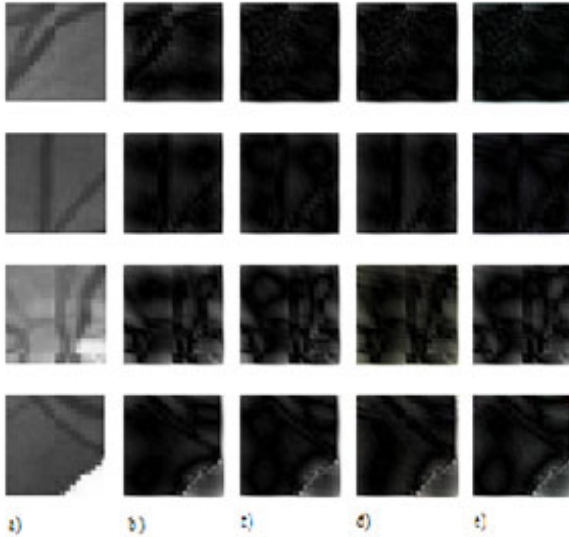


Fig. 9. Original block retina image (a) and real part of $Gab(x; y; \theta_i)$ for $\theta = 0$ (b), $\theta = 45$ (c), $\theta = 90$ (d), $\theta = 135$ (e)

Table 3. Mean, standard deviation and skewnes Gabor power some block (Fig. 9) retina image

Parameters	Retinal image 1		
	$\mu(\alpha, \sigma_x, \sigma_y)$	$std(\alpha, \sigma_x, \sigma_y)$	<i>Skew</i>
$\theta = 0$	25.971	33.468	3.712
$\theta = 45$	15.867	26.176	4.614
$\theta = 90$	15.900	26.901	4.525
$\theta = 135$	15.466	25.409	4.591
Parameters	Retinal image 2		
	$\mu(\alpha, \sigma_x, \sigma_y)$	$std(\alpha, \sigma_x, \sigma_y)$	<i>Skew</i>
$\theta = 0$	23.204	34.360	4.402
$\theta = 45$	18.494	32.176	4.655
$\theta = 90$	22.769	34.375	4.366
$\theta = 135$	16.307	27.644	4.781

3 Conclusion

A new method has been presented for iris and retina recognition based on Gabor and textural features. This paper analyses the details of the proposed method. Experimental results have demonstrated that this approach is promising to improve iris and retina recognition for person authentication.

References

1. Goh, K.G., Lee, M.L., Hsu, W., Wang, H.: ADRI: An Automatic Diabetic Retinal Image Screening System. In: Medical Data Mining and Knowledge Discovery, Springer, Heidelberg (2000)
2. Hsu, W., Pallawala, P.M.D.S., Lee, M.L., Kah-Guan, A.E.: The Role of Domain Knowledge in the Detection of Retinal Hard Exudates. In: IEEE Computer Vision and Pattern Recognition, Hawaii (December 2001)
3. Li, H., Chutatape, O.: Automated feature extraction in color retinal images by a model based approach. IEEE Trans. Biomed. Eng. 51, 246–254 (2004)
4. Salem, N.M., Nandi, A.K.: Novel and adaptive contribution of the red channel in pre-processing of colour fundus images. Journal of the Franklin Institute, 243–256 (2007)
5. Kirbas, C., Quek, K.: Vessel extraction techniques and algorithm: a survey. In: Proceedings of the 3rd IEEE Symposium on BioInformatics and BioEngineering, BIBE 2003 (2003)
6. Chang, S., Shim, D.: Sub-pixel Retinal Vessel Tracking and Measurement Using Modified Canny Edge Detection Method. Journal of Imaging Science and Technology (March-April 2008)
7. Chanwimaluang, T., Fan, G.: An efficient algorithm for extraction of anatomical structures in retinal images. In: Proc. IEEE International Conference on Image Processing, pp. 1093–1096 (2003)

8. Farzin, H., Abrishami-Moghaddam, H., Moin, M.S.: A novel retinal identification system. *EURASIP Journal on Advances in Signal Processing* 2008, Article ID 280635 (2008)
9. Choras, R.S.: Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems. *International Journal of Biology and Biomedical Engineering* 1(1), 6–16 (2007)
10. Choras, R.S.: Iris Recognition. In: Kurzynski, M., Wozniak, M. (eds.) *Computer Recognition Systems 3. AISC*, vol. 57, pp. 637–644. Springer, Heidelberg (2009)
11. Choras, R.S.: Iris-based person identification using Gabor wavelets and moments. In: *Proceedings 2009 International Conference on Biometrics and Kansei Engineering ICBAKE 2009*, pp. 55–59. CPS IEEE Computer Society, Los Alamitos (2009)
12. Gabor, D.: Theory of communication. *J. Inst. Elect. Eng.* 93, 429–459 (1946)
13. Haralick, R.M.: Statistical and structural approaches to texture. *IEEE Transaction on Systems, Man and Cybernetics* 67, 786–804 (1979)

Recognition and Classification

Fusion Methods for the Two Class Recognition Problem – Analytical and Experimental Results

Michał Woźniak and Marcin Zmysłony

Chair of Systems and Computer Networks
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: {Michał.Wozniak, Marcin.Zmyslony}@pwr.wroc.pl

Summary. In this paper we take into consideration group of decision making methods formed by the classifier fusion on the level of their discriminates. For such models we analyze what is the best way of assigning weights for them. Some analytical properties are of aforementioned methods are shown. Evaluation of proposed concept is done on the basis on computer experiment results.

1 Introduction

One of the main challenge for human brain is to find information in the given data effectively and classify unknown data as accurately as possible [1]. But the real problems sometimes are so complicate that one brain (classifier) is not strong enough to solve them. To improve the classification accuracy, multiple classifier systems (MCSs), could be used for practical tasks. The main idea of aforementioned proposition could be compared to a quite intuitive idea of a group of experts working together to solve a problem. It seems obvious that when dealing with complicated problem, group of experts, has higher probability of reaching satisfactory solution then a single expert.

When we consider MCSs we can treat the classifier outputs as the input to a second-level pattern recognition system and design a new classifier which is called *fuser*. There are several ways how to design it. The first one proposes that a fuser can make a decision on the basis of classifiers responses (class numbers given by individual classifiers) [3]. Initially one could find only the majority vote in literature, but in later works more advanced methods of receiving common answer of the classifier group were proposed. The second group of collective decision making methods is formed by the classifier fusion on the level of their discriminats.

In this paper we take into consideration the second group of methods and try to analyze what is the best way of assigning weights for them.

2 Fusion Based on Values of Classifiers Discriminant Function

The classification algorithm which is considered by us, is formed by the procedures of classifier fusions on the basis of their discriminants, the main form of which are *posterior* probability estimators, referring to the probabilistic model of a pattern recognition task [6], but we could propose a support function using by classifier to make its own decision e.g. neural nets output.

There are several propositions of designing such a fuser. First group of methods which do not require learning, like the aggregating methods, perform fusion with the help of simple operators like maximum or average. The problem is that such methods work well only in clearly defined conditions and that is the reason why they can not always be used in practical applications [5]. Alternative to the aggregating methods are weighting methods, where the selection of weights has a similar importance as it is in case of weighted majority voting.

In this model we assume that we have n classifiers. Each classifier makes decision based on the value of discriminant. Let $F^{(l)}(i, x)$ means such a function assigned to class i for given value of x , which is used by the l -th classifier $\Psi^{(l)}$. A common classifier $\Psi(\hat{x})$ looks as follows

$$\Psi(\hat{x}) = i \quad \text{if} \quad \hat{F}(i, x) = \max_{k \in M} \hat{F}(k, x), \quad (1)$$

were

$$\hat{F}(k, x) = \sum_{l=1}^n w^{(l)} F^{(l)}(i, x) \quad \text{and} \quad \sum_{l=1}^n w^{(l)} = 1 \quad (2)$$

In this model discriminant could be treated as the *posterior* probability for the classifiers based on Bayes decision theory or outputs of neural network. In general the value of such function means support given for distinguished class.

From the perspective of our recently work [7, 8], we claim that assigning weights in appropriate way is one of the main important factor of final decision of MCSs. That is why we analyze the problem of assigning weights in multi-aspects, which are as follows:

Weights dependent on classifier

This is a traditional approach where weights are connected with classifier and each discriminant function of given l -th classifier is weighted by the same value $w^{(l)}$. The estimation of probability error of such classifier could be found in [9].

Weights dependent on classifier and class number

Weight $w^{(l)}(i)$ is assigned to the l -th classifier and the i -th class. For given classifier assigned weights for different classes could be different.

Weights dependent on classifier and feature vector

Weight $w^{(l)}(x)$ is assigned to a given l -th classifier and for given x has the same value for each discriminant functions used by l -th classifier.

Weights dependent on classifier, class number, and feature vector

Weight $w^{(l)}(i, x)$ is assigned to the l -th classifier but for given x its value could be different for different discriminant function assigned to each classes.

The real challenge is to assign weights to model in such way, that even if the simple classifiers make wrong decision during the classification process, the final answer of classification system would be correct.

In next section we prove that it is possible to get correct final decision by fuser even if all simple classifiers pointed to the wrong class and we give conditions under which it is possible.

3 Analytical Characteristics of Fusion Methods

In this work we take into consideration two-class pattern recognition problem where answers of classifiers are multiply by weights dependent on classifier. In this case, as we mentioned before, weights are the same for all support functions given by particular classifier.

Let us note available decisions by $\{i, \bar{i}\}$ where i denotes correct class and \bar{i} wrong one. What is important, we assume that all simple classifiers point to wrong class. Base on this we can give a following theorem.

Theorem 1. *If for given x , all l -classifiers in committee point to wrong class, then there is no such a pool of $w^{(l)}(x)$ for which collective classifier would point to correct class, what means*

$$\exists_x \exists_{l\{1..n\}} \quad \text{if} \quad \Psi^{(l)}(x) = \bar{i} \quad \text{then} \quad \Rightarrow \hat{F}(i, x) < \hat{F}(\bar{i}, x).$$

Proof. For $w^{(l)}(x) \in [0,1]$ it is not possible to change of sign of aforementioned inequality by multiplying both sides by the same weights

$$w^{(l)}(x) * F^{(1)}(i|x) + \dots + w^{(n)}(x) * F^{(n)}(i|x) < w^{(1)}(x) * F^{(1)}(\bar{i}|x) + \dots + w^{(n)}(x) * F^{(n)}(\bar{i}|x)$$

because

$$\forall l w^{(l)}(x) * F^{(l)}(i|x) < w^{(l)}(x) * F^{(l)}(\bar{i}|x)$$

and

$$\forall l F^{(l)}(i|x) < F^{(l)}(\bar{i}|x),$$

because as we mentioned above each of individual classifier points on \bar{i} .

Of course creating fuser where weights dependent only on classifiers ($w^{(l)}$) gives also the same results, because it is special case of aforementioned model.

Let us consider similar classification problem and assume that all simple classifier points to wrong class \bar{i} , but let us consider a set of weights which dependent on classifier and class number. For such defined problem we present Theorem 2

Theorem 2. *If for particular x , all n classifiers in a pool point to wrong class, there is such pool of $w^{(l)}(x)$ for which combined classifier could point to correct class.*

Proof. For given individual classifier the support function for correct class is denoted as

$$\hat{F}(i|x) = \sum_{l=1}^L w^{(l)}(i) * F^l(i|x),$$

and for wrong class respectively

$$\hat{F}(\bar{i}|x) = \sum_{l=1}^L w^{(l)}(\bar{i}) * F^l(\bar{i}|x),$$

Because on one hand we assume that all simple classifiers are wrong and on the other hand we want to show that even in such situation it is possible to get correct final answer we could write inequality

$$\sum_{l=1}^L w^{(l)}(i) * F^l(i|x) > \sum_{l=1}^L w^{(l)}(\bar{i}) * F^l(\bar{i}|x).$$

As we mentioned in introduction, support function could be referring to the probabilistic model of a pattern recognition task. That is why we assumed two more assumptions. The first

$$F^{(l)}(i|x) = 1 - F^{(l)}(\bar{i}|x)$$

and the second one

$$w^{(l)}(i) + w^{(l)}(\bar{i}) = 1.$$

Based on previous assumptions we transform inequality to form

$$\sum_{l=1}^L w^{(l)}(i) * (1 - F^l(\bar{i}|x)) > \sum_{l=1}^L w^{(l)}(\bar{i}) * F^l(\bar{i}|x)$$

Next moving the component between two sides we get

$$\sum_{l=1}^L w^{(l)}(i) > \sum_{l=1}^L F^l(\bar{i}|x) * (w^{(l)}(i) + w^{(l)}(\bar{i}))$$

and finally

$$w^{(1)}(i) + \dots + w^{(n)}(i) > F^1(\bar{i}|x) + \dots + F^n(\bar{i}|x).$$

From the final form of the inequality we can see that it is possible to get correct final result even if all classifiers are wrong, if the sum of weights

which point to correct class is bigger then the sum of support functions for misclassified class.

Of course creating fuser where weights dependent on classifiers, class number and vector value give also the same results, because it is special case of aforementioned model.

It is worth noting that the Theorem 2 shows only possibility to get such a good results, therefore we should look for a fuser training method. An ensemble learning task leads to the problem how to establish the following vector W

$$W = [w^{(1)}, w^{(2)}, \dots, w^{(n)}]$$

which consists of weights assigned to each classifier (denoted as l) and each class number (denoted as i)

$$W^{(l)} = [w^{(l,i)}, w^{(l,i)}, \dots, w^{(l,i)}]^T.$$

We could formulate the following optimization problem. The weights should be established in such way to maximize the accuracy probability of fuser:

$$\Phi(W) = 1 - P_e(W),$$

where $P_e(W)$ is frequency of misclassification. Closer look at this optimization problem was presented in 2.

Trying to solve this optimization problem, there were prepared some experiments which will be presented in next section.

4 Experiments

The purpose of the experiments is to confirm the Theorem 1 and 2 from previous section. In experiments two kinds of fuser are taking into consideration:

- Fuser where weights depend on classifier (FC).
- Fuser where weights depend on classifier and class number (FCCN).

4.1 Set Up of Experiment

We used three databases generated from PRTools: "Highleyman", "Banana", and "Difficult". Each of them consisted of 300 examples and was described by two attributes and two classes. To provide diversity of simple classifiers that allows their local competences to be exploited, only slightly undertrained networks has been used (for which training process were early stopped). The details of used neural nets are as follow:

- 5 neurons in hidden layer,
- sigmoidal transfer function,

- back propagation learning algorithm,
- number of neurons in last layer equals number of classes of given experiment.

The fusion block was realized as genetic algorithm. Additionally the qualities of mentioned above classifiers were compared with *Oracle* classifier. Other set-up of experiments were as follow:

1. All experiments were carried out in Matlab environment using the PRtools toolbox [9] and own software.
2. Classifiers' errors were estimated using the ten fold cross validation method [6].

The set up of experiments for each database was different, that is why here are only presented average values. For GA the mutation function was Gaussian, population size equals 300, 8 members of elite, the fraction of genes swapped between individuals equals 0.8 and 300 iterations.

For each database the experiment was repeated 10 times with different parameter of population size for GA.

5 Results

The best results obtained in those experiments are presented in table below with additional information about result obtained by Oracle classifier and majority vote. Detailed results of tests are compared in Table 1.

Table 1. Comparison of experimental results for 3 databases

Database	Oracle	FC			FCCN		
		Min	Average	Max	Min	Average	Max
Difficult	0,128	0,302	0,402	0,521	0,082	0,175	0,201
Highleyman	0,191	0,201	0,228	0,306	0,146	0,278	0,373
Banana	0,026	0,107	0,122	0,144	0,218	0,291	0,362

5.1 Experimental Results Evaluation

The results show that situation presented in Theorem 1 and 2 are possible to obtain. According to them, there are clearly defined conditions, that for two-class problem only way to get correct result, even if all simple classifiers make wrong decision, is to assign weights dependent on classifier and class number.

As we mentioned in our previous papers [7, 8], one of the famous model, which is especially used for comparison experimental results, is *Oracle* classifier. The *Oracle* classifier point to correct class if any of individual classifier in committee, point to correct class. Because of the specify, *Oracle* is very

good model to compare with our experiments. As we can see in most any cases the results obtained by our model were better then the results obtain by *Oracle*. This also confirm the rightness of our assumptions.

In our opinion this fact give a new direction in searching some new way of assigning weights to a models. Although we realize that the scope of computer experiments were limited and it is still a lot of work to do in this field.

6 Final Remarks

Some methods of classifier fusion were discussed in this paper. For those methods we tried to analyze what is the best way of assigning weights for them. Based on aforementioned assumptions we presented two theorems. First of them shows that if there is fusion model where weights do not dependent on class number, then it is not possible to get correct result if all simple classifiers in committee are wrong. The second theorem shows possibility of giving correct answer by fuser even if all simple classifiers are wrong, but it is worth noting that it is possible only if weights at least dependent on class number. Evaluation of proposed concept was done on the basis on computer experiments results on computer generated databases.

Obtained results justify the use of weighted combination and they are similar as published in [10, 11, 12]. Unfortunately, as it was stated, it is not possible to determine weight values in the analytical way. However, it is hoped that in practical situations the weights can be appropriately set, either with the aid of a suitable expert, or else by applying the data training methods proposed earlier in the paper.

References

1. Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, San Francisco (2000)
2. Woźniak, M., Zmysłony, M.: Designing fuser on the basis of discriminants - evolutionary and neural methods of training. In: Corchado, E., Graña Romay, M., Manhaes Savio, A. (eds.) Hybrid Artificial Intelligence Systems. LNCS, vol. 6077, pp. 590–597. Springer, Heidelberg (2010)
3. Kuncheva, L.I.: Combining pattern classifiers: Methods and algorithms. Wiley, Chichester (2004)
4. Biggio, B., Fumera, G., Roli, F.: Bayesian Analysis of Linear Combiners. In: Haindl, M., Kittler, J., Roli, F. (eds.) MCS 2007. LNCS, vol. 4472, pp. 292–301. Springer, Heidelberg (2007)
5. Duin, R.P.W.: The Combining Classifier: to Train or Not to Train? In: Proc. of the ICPR 2002, Quebec City (2002)
6. Duda, R.O., et al.: Pattern Classification. Wiley-Interscience, Hoboken (2001)
7. Zmysłony, M., Woźniak, M., Jackowski, K.: Oracle - is it really the best possible model? In: Choraś, R.S., Zabłudowski, A. (eds.) Image Processing & Communications Challenges, pp. 112–119. Exit, Warsaw (2009)

8. Woźniak, M., Zmyślony, M.: Method of designing classifier fuser - evolutionary approach. Modelling and simulation of systems. In: Proceedings of the 44th Spring International Conference, pp. 88–92. MARQ, Ostrava (2010)
9. Woźniak, M.: Experiments on linear combiners. In: Pietka, E., Kawa, J. (eds.) Information technologies in biomedicine, pp. 445–452. Springer, Heidelberg (2008)
10. Duin, R.P.W., Tax, D.M.J.: Experiments with Classifier Combining Rules. In: Kittler, J., Roli, F. (eds.) MCS 2000. LNCS, vol. 1857, pp. 16–29. Springer, Heidelberg (2000)
11. Kuncheva, L.I., Bezdek, J.C., Duin, R.P.W.: Decision templates for multiple classifier fusion: an experimental comparison. *Pattern Recognition* 34, 299–314 (2001)
12. Van Erp, M., Vuurpijl, L.G., Schomaker, L.R.B.: An overview and comparison of voting methods for pattern recognition. In: Proc. of IWFHR 8, Canada, pp. 195–200 (2002)

Feature Type and Size Selection for AdaBoost Face Detection Algorithm

Jerzy Dembski

Faculty of Electronics, Telecommunications and Informatics,
Gdansk University of Technology, ul. Narutowicza 11/12,
80-952 Gdansk, Poland
e-mail: dembski@ue.eti.pg.gda.pl

Summary. The article presents different sets of Haar-like features defined for adaptive boosting (AdaBoost) algorithm for face detection. Apart from a simple set of pixel intensity differences between horizontally or vertically neighboring rectangles, the features based on rotated rectangles are considered. Additional parameter that limits the area on which the features are calculated is also introduced. The experiments carried out on the set of MIT 19×19 face and non-face examples showed the usefulness of particular types of features and their influence on generalization.

1 Introduction

The AdaBoost algorithm for face detection has become popular in recent 5 years after the publication of works by Viola and Jones [6, 7]. The main advantages of this method lie in the low computational complexity of the detection process, especially when the cascaded version is used, and versatility of the concept (it can be applied not only to faces but also to the detection of other objects). On the other hand, its main drawbacks are related to the brute-force style of training, which is the reason for enormous learning complexity and sometimes weak generalization.

During research described in this paper the method for reducing the necessary number of features and increasing the generalization of AdaBoost algorithm was proposed and verified experimentally. Some evolutionary methods helpful in searching for the best set of features were also suggested.

2 The AdaBoost Algorithm

The strong AdaBoost classifier is constructed from weak classifiers, which participate in the final decision proportionally to their weights (α). Therefore, the structure of AdaBoost classifier is similar to simple perceptron with

thresholded activation function, in which weak classifiers are treated as inputs. In comparison to other classifier committees, the novelty is due to the method of weak classifier selection. At the first step the weak classifier which classifies training examples with lowest error is chosen. After that the special weights (w) associated with particular examples are increased for badly classified examples. At the next training epoch the algorithm tries to add a weak classifier in order to reinforce the whole ensemble and to properly classify the examples that were misclassified at previous epoch. To achieve this the next weak classifier is chosen to minimize weighted training error sum. The training process continues till the classification error is equal to 0.

The training algorithm for the two-class problem is presented in the frame below:

```

for each training example  $\{(\mathbf{x}_1, d_1), (\mathbf{x}_2, d_2) \dots (\mathbf{x}_K, d_K)\}$ 
  initialize weights  $w_i = 1/K_c$ , where  $K_c$  is a number of examples which
  belong to the same class  $c$  as  $i$ -th example
  generate all possible weak classifiers  $h_j(\mathbf{x}), j = 1 \dots L$ 

for  $t = 1$  to  $T$  (number of weak classifiers to be selected)
  1. normalize weights:  $w_i \leftarrow \frac{w_i}{\sum_{j=1}^K w_j}$ 
  2. select the weak classifier  $h_t(\mathbf{x})$ , which minimizes the weighted classification error:  $\epsilon_t = \min_j \sum_{i=1}^K w_i |h_j(\mathbf{x}_i) - d_i|$ 
  3. decrease the weights of properly classified examples:  $w_i \leftarrow w_i \frac{\epsilon_t}{1 - \epsilon_t}$ 

endfor

```

The strong classification function $H(\mathbf{x})$ consisted with T weak classifiers is described by the formula:

$$H(\mathbf{x}) = \begin{cases} 1 & \text{if } \sum_{t=1}^T \alpha_t h_t(\mathbf{x}) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where $\alpha_t = \log(1 - \epsilon_t) - \log \epsilon_t$ is a weight of t -th weak classifier. There are some extensions of the basic AdaBoost algorithm to multiclass problems e.g. AdaBoost.M1 and Adaboost.M2 presented in details in [4, 5].

The main idea of using weak classifiers is that none of them must classify all the training examples properly; instead, it is enough that it provides a little decrease of the classification error after it is added to the previously selected classifiers. The simplicity or weakness can be treated as an advantage in this case because it enables fitting the strong classification structure to the given problem and at the same time provides good generalization.

3 AdaBoost Algorithm for Face Detection

The particular kind of weak classifier was proposed for the face detection system in [6, 7]. Each weak classifier is connected with a single feature and divides the linear feature space into two subspaces, each of which represents one class. Each weak classifier can be shown as a function $h(\mathbf{x}, f, \theta, p)$, where \mathbf{x} is an input image, f – selected feature, θ – partition threshold and p – polarity e.g the direction of class 1.

The subsequent stages of strong classifier construction for the set of two-feature examples are shown on Fig. 1. In the simplest case Haar wavelets are used as features in such a way that each feature is calculated as a difference of sums of pixel intensities between two neighboring rectangular regions. Each pair of rectangles can vary in vertical and horizontal size and can be placed in any possible location in the image window. For instance, in the case of 19×19 windows there are 34200 features like the one shown on Fig. 2a.

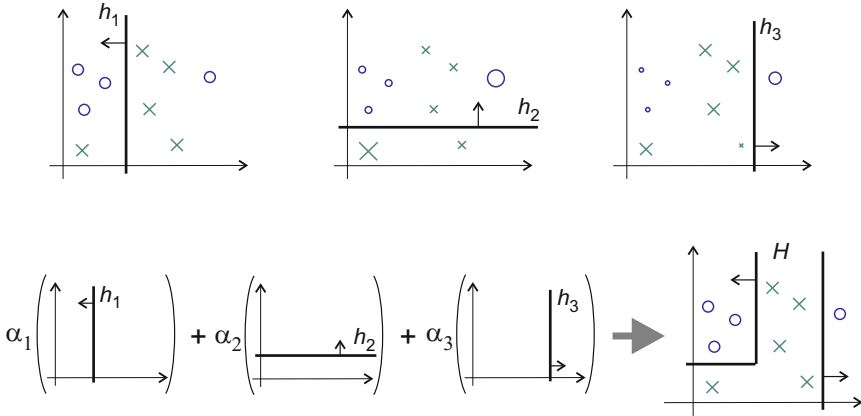


Fig. 1. AdaBoost training process with single-feature weak classifiers applied to a problem of two-feature vectors classification. The weights of particular samples are expressed with size of the icons (circles for samples from class 1, crosses for class 0), arrows indicate the direction of class 1. At the lower right corner strong classifier H is presented as a sum of decision boundaries.

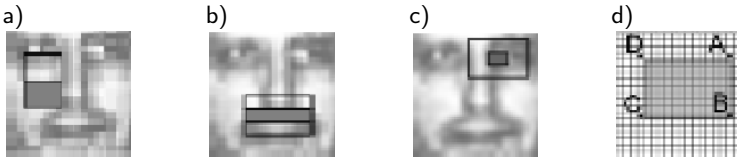


Fig. 2. Features based on intensity differences: a) edge, b) line, c) center-surround, d) calculation of the sum of pixel intensities using integral image

Two other features are shown on Fig. 2b,c. The neighborhood of edge and line features can be positioned vertically or horizontally. To avoid excessive computation complexity the integral image is calculated for each training example 7. Each entry of the integral array contains the sum of pixel intensities to the left and to the top of this entry. The sum of pixel intensities of ABCD rectangle presented on Fig. 2d can be calculated as a simple sum of four values: $Sum_{ABCD} = SAT(i_B, j_B) + SAT(i_D, j_D) - SAT(i_A, j_A) - SAT(i_C, j_C)$ where SAT denotes an integral array (*summed area table*).

Further complexity reduction can be reached by using the cascaded version of AdaBoost method 7. It is due to the possibility of threshold value manipulation in eq. (1). Lower threshold value increases the tolerance for detecting face in the image at the cost of increasing false positive error – many non-faces can be classified as faces. This is not a problem if strong classifier consists with many layers because each layer rejects non-faces more and more restrictively till the training error is equal to zero. The computational complexity reduction appears because of the two reasons: first, on a typical photograph there is a huge number of non-face windows and only a few actual faces; second, each layer in the cascade contains much less weak classifiers than the non-cascaded classifier, so each non-face is rejected very early by a quite simple part of the classifier.

4 Rotated Haar-Like Features

In 2 pixel intensity differences in rotated rectangle features were presented. However, due to some problems with the calculation of rotated integral images a new approach that consists of 3 stages is proposed here. At the first and the second stage the supporting array $R_{N \times M}$ is calculated by following directions shown on Fig. 3a,b; equation (2) is used at the first and (3) at the second stage. At the third stage the final array $RSAT$ (*rotated summed area table*) is created with eq. (4) by iterating from top to bottom and from left to right of R array.

$$R(i, j) = I(i, j) + R(i - 1, j - 1), \quad (2)$$

$$R(i, j) = R(i, j) + \begin{cases} R(i + 1, j - 1) & \text{if } i + 1 < N \\ R(i, j - 2) & \text{otherwise,} \end{cases} \quad (3)$$

$$RSAT(i, j) = R(i, j) + R(i, j - 1), \quad (4)$$

At each stage $R(i, j) = 0$ if $(i < 1)$ or $(j < 1)$ ¹. The sum of pixel intensities in rotated rectangular area is calculated in analogous way as in non-rotated case: $Sum_{ABCD} = RSAT(i_B, j_B) + RSAT(i_D, j_D) - RSAT(i_A, j_A) - RSAT(i_C, j_C)$. Fig. 4 presents three types of features which are based on rotated rectangles. The intensity differences are also computed in the same way as for non-rotated features (see section 3).

¹ Providing that rows and columns of arrays are indexed from 1.

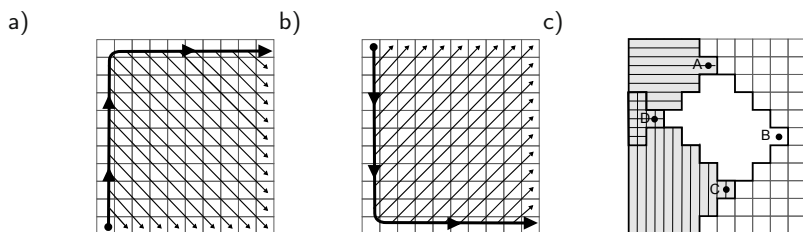


Fig. 3. Rotated integral image calculation: summing directions in outer loop (bold arrow) and inner loops (thin arrows); a) I stage, b) II stage, c) the sum of pixel intensities for ABCD rectangle

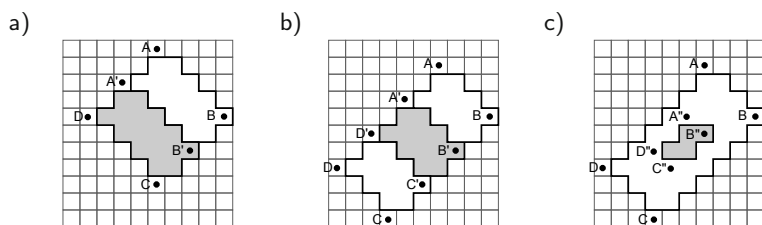


Fig. 4. Rotated rectangle features: a) edge, b) line, c) center-surround feature

It is important to multiply each sum by a factor that is inversely proportional to the area of the rectangle, in order to preserve the variance. For instance the feature at Fig. 4b can be calculated as $Sum_{ABCD} - 3Sum_{A'B'C'D'}$.

5 Experiments

Each experiment consisted of the training and the testing stage. For training we used a set of 6977 19×19 pixels examples (2429 faces and 4548 non-faces). The training process was considered complete when all training examples were correctly classified. For testing a separate set of 24045 examples (472 faces and 23573 non-faces) was used.

Fig. 5 presents the results by means of ROC curves. ROC shows the false negative classification error² with respect to the false positive error³ in the test set. We can move along the curve by changing the classification threshold, which standard value is equal to $\frac{1}{2} \sum_{t=1}^T \alpha_t$ as in eq. (11). Fig. 5a shows that the cascaded AdaBoost algorithm provides worse generalization than non-cascaded one but it may be caused by some problems with tolerance setting which in this case requires changes of threshold values in all layers;

² The rate of face images rejected as non-faces.

³ The rate of non-faces accepted as faces.

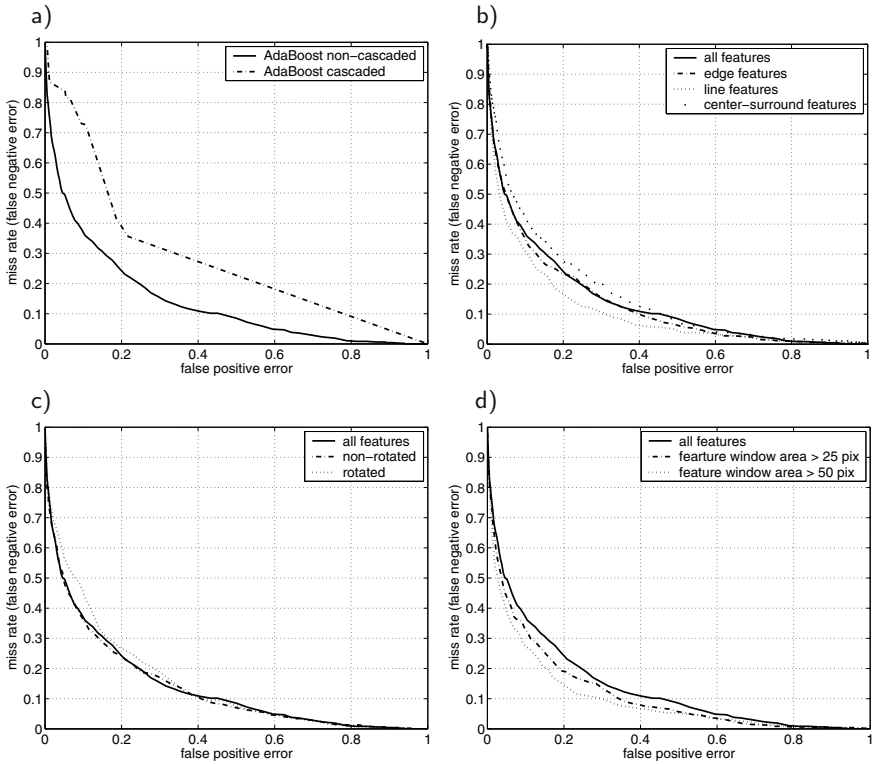


Fig. 5. Comparison of receiver operating characteristic (ROC) curves of the AdaBoost classifier: a) cascaded and non-cascaded training algorithm for all features, b) 3 different types of features, c) rotated and non-rotated rectangle features, d) introduction of additional limits of minimal feature rectangle area for all feature types

therefore, in subsequent experiments only non-cascaded algorithm was used. From Fig. 5b it can be seen that rotated and non-rotated line features are better than edge and center-surround features. On Fig. 5c rotated and non-rotated features are compared and it seems that non-rotated ones are slightly better. The most interesting results are related to Fig. 5d which shows that features built only from larger rectangles (for instance with area bigger than 50 pixels) increase generalization considerably.

6 Conclusions and Future Research

The general conclusion is that the experiments showed the influence of feature type and size on AdaBoost classifier generalization. The most spectacular increase of generalization was achieved by the little feature area reduction. It is probably due to the overfitting phenomenon known from machine learning

theory [3]. As a consequence of such remark, some kind of postprocessing of classification system seems to be required.

The generalization error in such a modified classifier training system can be computed using validation set created as a subset separated from the training set. The set of the best features selected firstly by AdaBoost algorithm in some training processes can be a base for the following genetic search. Two most promising ideas can be taken into account:

- each genetic individual (solution) relates to a feature subset and the fitness function is calculated as the correctness of the results obtained by the AdaBoost classifier trained on the particular feature subset,
- unlinear feature extraction [4] using genetic programming with some arithmetical functions serving as an alphabet and with features that were selected earlier treated as inputs.

Acknowledgements

This work was funded in part by national grant N N516 367936 and local research grant 018906.

References

1. Dembski, J., Jedruch, W.: Using similar classification tasks in feature extractor learning. In: Image Processing & Communications Challenges (IPC 2009), pp. 136–143. EXIT (2009)
2. Lienhart, R., Maydt, J.: An Extended Set of Haar-Like Features for Rapid Object Detection. In: IEEE ICIP 2002, pp. 900–903 (2002)
3. Mitchell, T.M.: Machine Learning. McGraw-Hill, New York (1997)
4. Schapire, R.E., Freund, Y.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 5591, 119–139 (1997)
5. Schapire, R.E., Freund, Y.: Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods. The Annals of Statistics 26(5), 1651–1686 (1998)
6. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features, pp. 511–518 (2001)
7. Viola, P., Jones, M.: Robust Real-Time Face Detection. International Journal of Computer Vision 57(2), 137–154 (2004)

3D Morphable Models Application for Expanding Face Database Limited to Single Frontal Face Image Per Person

Lukasz Kulasek and Andrzej Czyżewski

Gdansk University of Technology, Multimedia Systems Departament,
Electronics, Informatics and Telecommunications Faculty,
ul. Narutowicza 11/12 80-233 Gdańsk, Poland
e-mail: ksm@sound.eti.pg.gda.pl

Summary. Major problem in real world scenarios is lack of sufficient image samples per person in database for successful face recognition. In most cases insufficient number of samples per an individual in database is present. This makes face classification almost impossible for larger number of people. This problem is commonly described as 'one sample problem'. Recent state-of-art in face recognition allows to achieve high accuracy using face images with frontal pose. However, recognizing faces with rotations in depth, increases error rate significantly. In this paper we present a method to expand database using 3D morphable models to reconstruct 3D face from a single frontal image sample. By rotating reconstructed face to different views we create series of novel virtual images with pose variations for every individual in database. This approach can help to decrease error rate from pose variations and resolves 'one sample problem'.

1 Introduction

Despite over 30 years of research, face recognition still remains unsolved problem. One of the biggest issue in real world scenarios is lack of sufficient amount of image samples per person. This implicates that for the individual in database not enough image samples are available with pose variations. Many studies shown the decrease performance in face recognition with pose rotations in depth [6]. Thus, the need of expanding database with face images samples is crucial.

In the past, different attempts to solve this problem were made. In some of them generation of novel virtual views were done by using parallel deformation [1] or combining facial regions into bunches and shifting their position [3]. Despite computational incomplexity of these methods, they proved to solve problem with out of plane pose rotations.

Recently decrease of computational and memory storage cost makes 3D approach more available to apply in face recognition algorithms. Some methods use two orthogonal mugshot images to reconstruct 3D geometry and texture [9], while methods based on 3D morphable models need only one sample. These methods allow to construct new samples of face under different pose [7] or lightning [8]. Some of approaches model expression by application of MPEG-4 facial animations parameters [2]. Recent methods allow minimizing aging problem [4].

In this work we present a method based on 3D morphable model to reconstruct face geometry using basic mathematical calculations with fair accuracy.

2 Model Construction

2.1 3D Face Scanner

3D scanner based on Phase Shifting Interferometry method (PSI) was constructed to acquire 3D face scans needed for construction of morphable models. 29 scans of women faces and 25 of men faces were made. Depth information and texture were calculated and saved in separate files. Depending on resolution, each file contains (x, y, z) coordinates of 60000 to 100000 points in three dimensional space, its (r, g, b) color values and *flag field* for each point. These fields store a number from range $(0, 255)$ and are used to include additional information. In this research we used these fields to store a number of facial point to which given vertex belongs during creation of morphable models.

2.2 Generic Face Model Construction

Three generic models were created from face scans divided into three groups as seen in Table 1. First group contained male scans, second female scans and third contained all face scans together. First two groups were used to create two morphable models to reconstruct 3D face geometry of given individual depending of his or her gender. Third group was used to create morphable model for databases where gender information is not available, or for comparative studies.

Table 1. Number of 3D face scans for each constructed morphable model

Morphable model type	Number of 3D face scans
male	25
female	29
male+female	54

In each 3D face scan set of 13 facial positions μ corresponding to eye corners, nose tip, lips etc. were manually selected. The order of these points is shown in Figure 1.

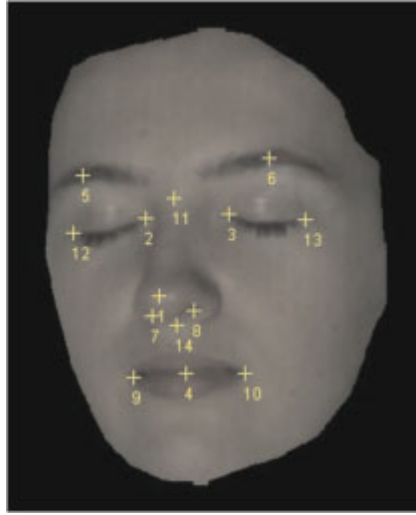


Fig. 1. Order of assigning flags to facial points in one of 3D face scans

These positions are described as

$$\mu = (\mu_1, \mu_2, \dots, \mu_{13}) \quad (1)$$

where μ_n is (x, y, z) coordinates of n -th facial point in face scan.

In each group, faces were aligned and normalised according to the positions of facial points. Depth and texture values were averaged. Three generic models shown in Figure 2 were created.

3 Generation of Novel Virtual Face Samples

3.1 Fitting Morphable Model

In each created morphable model new facial points $\beta = (\beta_1, \beta_2, \dots, \beta_{13})$ were manually selected, as well as corresponding points $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_{13})$ in input image.

Fitting algorithm makes a number of iterations, rotates, scales and translates positions of β using least squares method. Given two sets of facial points α and β we find minimum of

$$S = \sum_{n=1}^{13} (\alpha_n - \beta_n)^2 \quad (2)$$

where α_n and β_n are (x, y) coordinates of n -th facial point in image and morphable model respectively.



Fig. 2. 3 generic models created: male model (left), female model (middle) and male+female model (right)

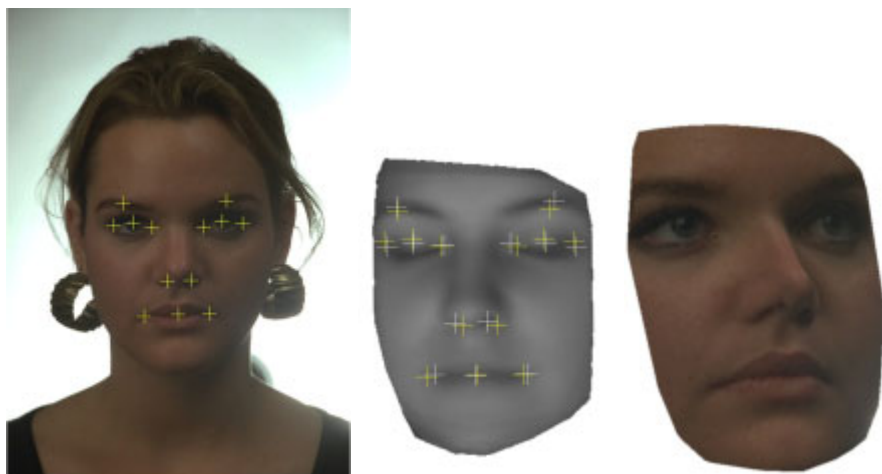


Fig. 3. 3D face reconstruction from sample image [5] (left), by fitting morphable model (middle) and geometry and texture in result (right)

When minimum is found, α and β are in best match. Texture is transferred to morphable model and new 3D face geometry is created as shown in Fig. 3.

3.2 Performance

We tested the algorithm on 200 face images from FERET database [5]. A comparison of created samples with real images are shown in Fig. 4. Virtual samples in range of pose variations are shown in Fig. 5. In most cases face reconstruction results were fair. In some, errors in texture positioning occurred near nose tip and nostrils areas. This suggests choosing facial points position on edges or corners in future work. Some quality loss could be seen for larger rotations.



Fig. 4. Performance of proposed algorithm compared with real images from FERET database [5]. Columns from left: input image, created virtual sample (middle) and real image with pose variation (right).



Fig. 5. Frontal face image from FERET database [5] (left) and constructed corresponding novel samples

4 Conclusions

Given only one frontal face image we can create any number of additional samples with any pose rotation. Use of incomplex mathematical calculations makes proposed method fast and working with a fair accuracy.

In future work we plan to implement facial points detector to automate texture transfer. Geometry morphing algorithm will be implemented to make reconstruction more accurate.

Acknowledgments

Research funded within the project No. POIG.02.03.03-00-008/08, entitled "MAYDAY EURO 2012- the supercomputer platform of context-depended analysis of multimedia data streams for identifying specified objects or safety threads". The project is subsidized by the European regional development fund and by the Polish State budget.

References

1. Beymer, D., Poggio, T.: Face Recognition From One Example View, A.I. Memo No. 1536, C.B.C.L. Paper No. 121 (September 1995)
2. Jiang, D., Hu, Y., Yan, S., Zhang, L., Zhang, H., Gao, W.: Efficient 3D Reconstruction for Face Recognition, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China (2005)
3. Huang, J., Yuen, P.C., Chen, W., Lai, J.H.: Component-based LDA Method for Face Recognition with One Training Sample. In: Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures, AMFG (2003)

4. Unsang, P., Yiying, T., Anil, J.K.: Face Recognition with Temporal Invariance: A 3D Aging Model. Michigan State University, East Lansing (2008)
5. Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office and citations to: *a.* Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.: The FERET database and evaluation procedure for face recognition algorithms. *Image and Vision Computing J* 16(5), 295–306 (1998); *b.* Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET Evaluation Methodology for Face Recognition Algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence* 22, 1090–1104 (2000)
6. Xiaoyang, T., Songcan, C., Zhi-Hua, Z., Fuyan, Z.: Face Recognition from a Single Image per Person: A Survey, Department of Computer Science and Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing, China (2006)
7. Bai, X., Yin, B., Shi, Q., Sun, Y.: Face Recognition Using Extended Fisherface With 3D Morphable Model. In: Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, August 18-21 (2005)
8. Zhang, X., Gao, Y., Leung, M.K.H.: Automatic Texture Synthesis for Face Recognition from Single Views. In: 18th International Conference on Pattern Recognition (2006)
9. Xiaozheng, Z., Yongsheng, G., Bai-ling, Z.: Recognizing Rotated Faces from Two Orthogonal Views in Mugshot Databases. In: Proceedings of the 18th International Conference on Pattern Recognition (2006)

A Partition of Feature Space Based on Information Energy in Classification with Fuzzy Observations

Robert Burduk

Department of Systems and Computer Networks
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: robert.burduk@pwr.wroc.pl

Summary. The paper considers the partition problem of feature space in classification. The partition is based on information energy for fuzzy events. In this paper we use Bayes rule for classification with fuzzy observations and exact classes. Additionally a probability of misclassifications is derived for fuzzy information on object features. The results show deterioration the quality of classification when we use fuzzy information on object features instead of exact information and are compared with the partition of feature space. Numerical example concludes the work.

1 Introduction

The choice of a partition of feature space is crucial to the performance of a classification system. For this systems the classification error is the ultimate measure of the performance of a classifier. Several studies have previously described the Bayes probability of error for a single-stage classifier [1], [2], for a combining classifiers [14] and for a hierarchical classifier [6], [8]. Some studies pertaining to bounds on the probability of error in fuzzy concept are given in [3], [9], [12], [11]. This paper considers the problem of the probability of error in a Bayes classifier when observations of the features are represented by the fuzzy sets. The recited results are compared with the partitions of feature space based on information energy of fuzzy events.

There are many cases when the available information is a mixture of randomness and fuzziness. In [9] formulated the pattern recognition problem with fuzzy classes and fuzzy information and consider the following tree situations:

- fuzzy classes and exact information,
- exact classes and fuzzy information,
- fuzzy classes and fuzzy information.

In this paper, our aim is to present the partition of feature space based on information energy in classification with fuzzy observations in context of the probability of error. This error is calculated for Bayes rule applied to pattern recognition with fuzzy observations. In this situation we use exact classes and fuzzy information of object features.

The contents of the work are as follows. Section 2 introduces the necessary background and describes the Bayes classifier. In section 3 the basic notions of fuzzy theory and the information energy of fuzzy event are presented. In section 4 we presented the difference between the probability of misclassification for the fuzzy and crisp data in Bayes optimal classifier. In numerical example the information energy and difference between the probability of misclassification are compared.

2 Bayes Classifier

Bayesian decision theory is a fundamental statistical approach to the problem of pattern classification. This approach is based on quantifying the tradeoffs between various classification decision using probability and the costs that accompany such decision. It makes the assumption that the decision problem is posed in probabilistic terms, and that all of the probability values are known.

A pattern is represented by a set of d features, or attributes, viewed as a d -dimensional feature vector $x \in \mathfrak{R}^d$.

Let us consider a pattern recognition problem, in which the class label ω is a random variable taking values in the set of class labels $\Omega = \{\omega_1, \dots, \omega_c\}$. The *priori probabilities*, $P(\omega_i)$, $i = 1, \dots, c$ constitute the probability mass function of the variable ω , $\sum_{i=1}^c P(\omega_i) = 1$. Assume that the objects from class ω_i are distributed in $x \in \mathfrak{R}^d$ according to the *class-conditional probability density function* $p(x|\omega_i)$, $p(x|\omega_i) \geq 0$, $\forall x \in \mathfrak{R}^d$, and $\int_{\mathfrak{R}^d} p(x|\omega_i)dx = 1$, $i = 1, \dots, c$.

Given the prior probabilities and the *class-conditional probability density functions* we can calculate the *posterior probability* that the true class label of the measured x is ω_i using the Bayes formula

$$P(\omega_i|x) = \frac{P(\omega_i)p(x|\omega_i)}{p(x)} \quad (1)$$

where $p(x) = \sum_{i=1}^c P(\omega_i)p(x|\omega_i)$ is the unconditional likelihood of $x \in \mathfrak{R}^d$.

Equation (1) gives the probability mass function of the class label variable ω for the observed x . The decision for that particular x should be made with respect to the posterior probability.

The "optimal" Bayes decision rule for minimizing the risk (expected value of the loss function) can be stated as follows: Assign input pattern x to class ω_i for which the conditional risk

$$R^*(\omega_i|x) = \sum_{j=1}^c L(\omega_i, \omega_j)P(\omega_j|x) \quad (2)$$

is minimum, where $L(\omega_i, \omega_j)$ is the loss incurred in deciding ω_i when the true class is ω_j . The Bayes risk, denoted R^* , is the best performance that can be achieved. In the case of the zero-one loss function

$$L(\omega_i, \omega_j) = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases},$$

the conditional risk becomes the conditional probability of misclassification and optimal Bayes decision rule is as follows:

$$R^*(\omega_i|x) = \max_i P(\omega_i|x). \tag{3}$$

Let Ψ^* be a classifier that always assigns the class label with the largest posterior probability. The classifier based on Bayes rule is the following:

$$\Psi^*(x) = \omega_i \quad \text{if} \quad \omega_i = \arg \max_i P(\omega_i)p(x|\omega_i). \tag{4}$$

because the unconditional likelihood $p(x) = \sum_{i=1}^c P(\omega_i)p(x|\omega_i)$ is even for every class ω_i

2.1 Bayes Error

The error of Ψ^* is the smallest possible error, called the Bayes error. The overall probability of error of Ψ^* is the sum of the errors of the individual x s weighted by their likelihood values $p(x)$,

$$Pe(\Psi^*) = \int_{\mathfrak{R}^d} [1 - P(\omega_i^*|x)]p(x)dx. \tag{5}$$

It is convenient to split the integral into c integrals, one on each classification region. For this case class ω_i^* will be specified by the region's label. Then

$$Pe(\Psi^*) = \sum_{i=1}^c \int_{\mathfrak{R}_i^*} [1 - P(\omega_i|x)]p(x)dx \tag{6}$$

where \mathfrak{R}_i^* is the classification region for class ω_i , $\mathfrak{R}_i^* \cap \mathfrak{R}_j^* = 0$ for any $i \neq j$ and $\bigcup_{i=1}^c \mathfrak{R}_i^* = \mathfrak{R}^d$. Substituting (5) into (6) we have (7):

$$Pe(\Psi^*) = 1 - \sum_{i=1}^c \int_{\mathfrak{R}_i^*} P(\omega_i)p(x|\omega_i)dx. \tag{7}$$

In Fig. 1 the Bayes error is presented for the simple case of $x \in \mathfrak{R}$, $\Omega = \{\omega_1, \omega_2\}$ and $P(\omega_1|x) = 1 - P(\omega_2|x)$. According to (6) the Bayes error is the area under $P(\omega_2)p(x|\omega_2)$ in \mathfrak{R}_1^* plus the area under $P(\omega_1)p(x|\omega_1)$ in \mathfrak{R}_2^* . The total area corresponding to the Bayes error is marked in black.

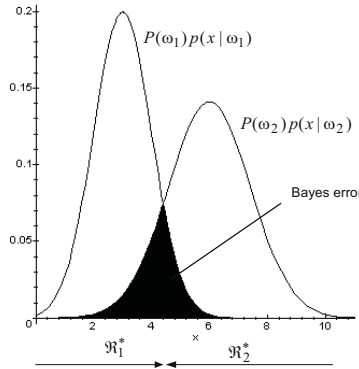


Fig. 1. The probability of error for Bayes optimal classifier when object features are non-fuzzy

3 Basic Notions of Fuzzy Theory

Fuzzy number A is a fuzzy set defined on the set of real numbers \mathbb{R} characterized by means of a membership function $\mu_A(x)$, $\mu_A : \mathbb{R} \rightarrow [0, 1]$:

$$\mu_A(x) = \begin{cases} 0 & \text{for } x \leq a, \\ f_A(x) & \text{for } a \leq x \leq c, \\ 1 & \text{for } c \leq x \leq d, \\ g_A(x) & \text{for } d \leq x \leq b, \\ 0 & \text{for } x \geq b, \end{cases}$$

where f_A and g_A are continuous functions, f_A is increasing (from 0 to 1), g_A is decreasing (from 1 to 0). In special cases it may be $a = -\infty$ and (or) $b = +\infty$. In this study, the special kinds of fuzzy numbers including triangular fuzzy numbers is employed. A triangular fuzzy numbers can be defined by a triplet $A = (a_1, a_2, a_3)$. The membership function is

$$\mu_A(x) = \begin{cases} 0 & \text{for } x \leq a_1, \\ (x - a_1)/(a_2 - a_1) & \text{for } a_1 \leq x \leq a_2, \\ (a_3 - x)/(a_3 - a_2) & \text{for } a_2 \leq x \leq a_3, \\ 0 & \text{for } x \geq a_3. \end{cases}$$

The width w_A of the fuzzy number A is defined as following value [4]:

$$w_A = \int_{-\infty}^{+\infty} \mu_A(x) dx. \tag{8}$$

A fuzzy information $\mathcal{A}_k \in \mathfrak{R}^d$, $k = 1, \dots, d$ (d is the dimension of the feature vector) is a set of fuzzy events $\mathcal{A}_k = \{A_k^1, A_k^2, \dots, A_k^{n_k}\}$ characterized by membership functions

$$\mathcal{A}_k = \{\mu_{A_k^1}(x_k), \mu_{A_k^2}(x_k), \dots, \mu_{A_k^{n_k}}(x_k)\}. \tag{9}$$

The value of index n_k defines the possible number of fuzzy events for x_k (for the k -th dimension of feature vector). In addition, assume that for each observation subspace x_k the set of all available fuzzy observations (9) satisfies the orthogonality constraint (13):

$$\sum_{l=1}^{n_k} \mu_{A_k^l}(x_k) = 1. \tag{10}$$

The probability of fuzzy event assume in Zadeh’s form (15):

$$P(A) = \int_{\mathbb{R}^d} \mu_A(x) f(x) dx. \tag{11}$$

The probability $P(A)$ of a fuzzy event A defined by (11) represents a crisp number in the interval $[0, 1]$.

The information energy contained in the fuzzy event A is defined by (10):

$$W(A) = P(A)^2 + P(\bar{A})^2, \tag{12}$$

where $P(\bar{A})$ is the complement set of A .

The information energy contained in the fuzzy information \mathcal{A} is defined by (10):

$$W(\mathcal{A}) = \sum_{l=1}^k P(A_l)^2. \tag{13}$$

4 Probability Error in Bayes Classifier with Fuzzy Observations

When we have non-fuzzy observation of object features in Bayes classifier then recognition algorithm for zero-one loss function is given by (3) and probability of error is given by (6). Similarly, if (10) holds and we use probability of fuzzy event given by (11) the Bayes recognition algorithm for fuzzy observations \mathcal{A} is the following:

$$\Psi_F^*(\tilde{A}) = \omega_i \quad \text{if} \tag{14}$$

$$\omega_i = \arg \max_i P(\omega_i) \int_{\mathbb{R}^d} \mu_{\tilde{A}}(x) p(x|\omega_i).$$

The probability of error $Pe(\Psi_F^*)$ for fuzzy data is the following:

$$Pe(\Psi_F^*) = 1 - \sum_{i=1}^c \sum_{\tilde{A} \in i} P(\omega_i) \int_{\mathbb{R}_i^*} \mu_{\tilde{A}}(x) p(x|\omega_i) dx, \tag{15}$$

where $\tilde{A} \in i$ denote the fuzzy observations belongs to the i -th classification region.

When we use fuzzy information on object features instead of exact information we deteriorate the classification accuracy. The difference between the probability of misclassification for the fuzzy $Pe(\Psi_F^*)$ and crisp data $Pe(\Psi^*)$ in Bayes optimal classifier is the following:

$$\begin{aligned}
 & Pe(\Psi_F^*) - Pe(\Psi^*) = \tag{16} \\
 & = \sum_{\tilde{A} \in \mathbb{R}^d} \left(\int_{\mathbb{R}^d} \mu_{\tilde{A}}(x) \max_i \{P(\omega_i)p(x|\omega_i)\} dx - \right. \\
 & \quad \left. - \max_i \left\{ \int_{\mathbb{R}^d} \mu_{\tilde{A}}(x) P(\omega_i)p(x|\omega_i) dx \right\} \right).
 \end{aligned}$$

similarly as in [3].

The element of $\sum_{\tilde{A} \in \mathbb{R}^d}$ equals to 0 if and only if, for the support of fuzzy observation \tilde{A} , one of the i discriminant functions $[P(\omega_1)p(x|\omega_1), \dots, P(\omega_i)p(x|\omega_i)]$ is uniformly larger than the others. Another interpretation is that the value of equation (16) depends only from these the observation, in whose supports intersect the discriminant functions.

4.1 Numerical Example

Let us consider the binary classifier with *a priori* probabilities $P(\omega_1) = P(\omega_2) = 0.5$. The class-conditional probability density functions are normal distributions in \mathbb{R}^1 : $p(x|\omega_1) = N(1, 1)$ and $p(x|\omega_2) = N(2, 1)$.

In experiments, the following sets of fuzzy numbers were used:
 case A

$$\mathcal{A} = \{A^1 = (-\infty, 0, 1), A^2 = (0, 1, 2), \dots, A^5 = (3, 4, \infty)\},$$

case B

$$\mathcal{B} = \{B^1 = (-\infty, 0, 1), B^2 = (0, 1, 2), \dots, B^5 = (3, 4, 5), B^6 = (4, 5, \infty)\},$$

Tab. 1 and 2 show the information energy and the difference between the probability of misclassification for fuzzy and non fuzzy data in the Bayes optimal classification for case A and B respectively. The information energy is calculated from (13) and the difference in probability of misclassification from (16). These results are calculated for a full probabilistic information.

In two cases we have different partitions of feature space. In each cases the width of fuzzy events are the same. The received results show the best position

Table 1. The information energy $W(\mathcal{A})$ and difference between the probability of misclassification $Pe(\Psi_F^*) - Pe(\Psi^*)$ in case A

	$p((x - k) \omega_1),$				$p((x - k) \omega_2),$		$k =$		
	0	0.25	0.5	0.75	1	1.25	1.5	1.75	2
$W(\mathcal{A})$	0.249	0.243	0.240	0.246	0.249	0.263	0.284	0.320	0.367
$Pe(\Psi_F^*) - Pe(\Psi^*)$	0.024	0.010	0.005	0.010	0.024	0.010	0.005	0.010	0.024

Table 2. The information energy $W(\mathcal{B})$ and difference between the probability of misclassification $Pe(\Psi_F^*) - Pe(\Psi^*)$ in case B

	$p((x - k) \omega_1),$				$p((x - k) \omega_2),$		$k =$		
	0	0.25	0.5	0.75	1	1.25	1.5	1.75	2
$W(\mathcal{B})$	0.249	0.242	0.238	0.237	0.236	0.237	0.238	0.242	0.249
$Pe(\Psi_F^*) - Pe(\Psi^*)$	0.024	0.010	0.005	0.010	0.024	0.010	0.005	0.010	0.024

of the class-conditional probability density functions in this partition. This position is appointed by the minimum of information energy.

The received results show also deterioration the quality of classification when we use fuzzy information on object features instead of exact information in Bayes optimal classifier. We have to notice that the difference in the misclassification for fuzzy and crisp data does not depend only on the fuzzy number. The parameter k shifts the discriminant functions $P(\omega_1)p((x - k)|\omega_1)$ and $P(\omega_2)p((x - k)|\omega_2)$. The position of the class-conditional probability density in relation to the observed fuzzy features is the essential influence for the difference $Pe(\Psi_F^*) - Pe(\Psi^*)$. The minimum of difference between the probability of misclassification is not always received for the smallest value of information energy.

5 Conclusion

In the present paper we have concentrated on the Bayes optimal classifier. Assuming a full probabilistic information we have presented the difference between the probability of misclassification for fuzzy and crisp data. Additionally, the partitions of feature space based on information energy of fuzzy events are presented. The obtained results show that the smallest difference between the probability of misclassification for fuzzy and crisp data and the partitions of feature space based on information energy of fuzzy event are independent.

Acknowledgements

This work is supported in part by The Polish State Committee for Scientific Research under the grant which is being realized in years 2010–2013.

References

1. Antos, A., Devroye, L., Györfi, L.: Lower Bounds for Bayes Error Estimation. *IEEE Trans. Pattern Analysis and Machine Intelligence* 21, 643–645 (1999)
2. Avi-Itzhak, H., Diep, T.: Arbitrarily Tight Upper and Lower Bounds on the Bayesian Probability of Error. *IEEE Trans. Pattern Analysis and Machine Intelligence* 18, 89–91 (1996)
3. Burduk, R.: Classification error in Bayes multistage recognition task with fuzzy observations. *Pattern Analysis and Applications* 13(1), 85–91 (2010)
4. Hanas, S.: On the Interval Approximation of a Fuzzy Number. *Fuzzy Sets and Systems* 122, 353–356 (2001)
5. Janickow, C.Z.: Fuzzy Decision Trees: Issues and Methods. *IEEE Trans. Systems, Man, and Cybernetics B: Cybernetics* 28, 1–14 (1998)
6. Kulkarni, A.: On the Mean Accuracy of Hierarchical Classifiers. *IEEE Transactions on Computers* 27, 771–776 (1978)
7. Kuncheva, L.I.: *Combining Pattern Classifier: Methods and Algorithms*. John Wiley, New York (2004)
8. Kurzyński, M.: On the Multistage Bayes Classifier. *Pattern Recognition* 21, 355–365 (1988)
9. Okuda, T., Tanaka, H., Asai, K.: A formulation of fuzzy decision problems with fuzzy information using probability measures of fuzzy events. *Information and Control* 38, 135–147 (1978)
10. Pardo, L.: Information Energy of a Fuzzy Event and a Fuzzy Events. *IEEE Trans. on Systems, Man, and Cybernetics SMC-15*(1), 139–144 (1985)
11. Pardo, J.A., Taneja, I.J.: On the Probability of Error in Fuzzy discrimination Problems. *Kybernetes* 21(6), 43–52 (1992)
12. Pardo, L., Menendez, M.L.: Some Bounds on Probability of Error in Fuzzy Discrimination Problems. *European Journal of Operational Research* 53, 362–370 (1991)
13. Pedrycz, W.: Fuzzy Sets in Pattern Recognition: Methodology and Methods. *Pattern Recognition* 23, 121–146 (1990)
14. Woźniak, M.: Experiments on linear combiners. *Advances in Soft Computing* 47, 445–452 (2008)
15. Zadeh, L.A.: Probability Measures of Fuzzy Events. *Journal of Mathematical Analysis and Applications* 23, 421–427 (1968)

Recognition of Signed Expressions Using Cluster-Based Segmentation of Time Series

Mariusz Oszust and Marian Wysocki

Rzeszow University of Technology
Department of Computer and Control Engineering
W. Pola 2, 35-959 Rzeszow, Poland
e-mail: {moszust,mwysocki}@prz-rzeszow.pl

Summary. The paper considers automatic visual recognition of signed expressions. The proposed method is based on modeling gestures with subunits, which is similar to modeling speech by means of phonemes. To define the subunits a data-driven procedure is applied. The procedure consists in partitioning time series, extracted from video, into subsequences which form homogeneous groups. The cut points are determined by an evolutionary optimization procedure based on multicriteria quality assessment of the resulting clusters. In the paper the problem is formulated, its solution method is proposed and experimentally verified on a database of 100 Polish words.

1 Introduction

Automatic vision-based sign language recognition is an important prospective application of gesture-based human-computer interfaces. The aim of the research is a system that properly interprets gestures, e.g. translates them into written or spoken language. Most of such systems described in the literature (see e.g. [1], [2]) are based on word models where one sign represents one model in the model database. They can achieve good performance only with small vocabularies or gesture data sets. The training corpus and the training complexity increase with vocabulary size. So, large-vocabulary systems require the modeling of signed expressions in smaller units than words i.e. the words are modeled with subunits, which is similar to modeling speech by means of phonemes. The main advantage of this approach is that an enlargement of the vocabulary can be achieved by composing new signs through concatenation of subunit models and by tuning the composite model with only small sets of examples. However, an additional knowledge of how to break down signs into subunits is needed.

Different vision-based subunit segmentation algorithms have been developed. Following Liddell and Johnson's movement-hold model the authors of [3] propose modeling each sign (word) as a series of movement and hold

segments. Kraiss et al. in [1] present an iterative process of data-driven extraction of subunits using hidden Markov models (HMMs). Han et al. in [4] define the subunit boundary using hand motion discontinuity.

In this paper we propose a new approach where the subunits' boundary points are considered as decision variables in a multiobjective optimization problem. The problem consists in finding subunits which can be grouped in clusters of good quality. The quality is measured by two cluster validity indices, one based on entropy [5] and another the Dunn index [6, 7]. The indices are optimized simultaneously using lexicographic ordering [8] and an immune-based evolutionary algorithm [9, 10]. The approach refers to clustering of time series data [11, 12], multiobjective clustering [13, 14], and cluster-based time series segmentation [15]. The contribution of the paper lies in (1) formulation of the problem of determining subunits for sign language recognition as a multiobjective cluster optimization, (2) formulation of the problem of modeling signed expressions with the subunits, (3) proposition of solution methods, and verification of the approach by experiments on real data.

The rest of the paper is organized as follows. Section 2 formulates the problem of subunits extraction and describes the solution method. Section 3 gives details of the subunit-based recognition method. The results of experiments with recognition of 100 isolated words of the Polish Sign Language (PSL) are given in section 4. Section 5 concludes the paper.

2 A Data-Driven Subsequence Extraction Method

2.1 The Input Data

Most of the sign gestures are two-handed and dynamic. Let $S = \{X_1, X_2, \dots, X_n\}$ denote a data set, where a sequence $X_i = \{x_i(1), x_i(2), \dots, x_i(T_i)\}$ of real valued feature vectors represents a signed word. All vectors $x_i(t)$, where $i \in I = \{1, 2, \dots, n\}$, and t is a time sampling point, $t \in T_i = \{1, 2, \dots, T_i\}$, have identical structure. They contain features extracted from image sequences registered by a camera. For instance, we use seven manual features for the right hand and the same features for the left hand: the position of the hand with respect to the face (three spatial coordinates), the area, orientation, compactness and eccentricity of a hand (four features as a very simplified information about the hand shape). Two time sequences X_i and $X_{j \neq i}$ may represent different words or different realizations of the same word. In modeling signed expressions we should take into account that the features we are observing appear both sequentially and simultaneously. For example, the hand shape and hand position can change independently at the same time [1]. To model parallel processes we will distinguish N groups of features (channels). This is based on the assumption that the separate processes evolve independently from one another with independent output. So, we will write $x_i(t) = [x_i^1(t), x_i^2(t), \dots, x_i^N(t)]$ and, in accordance

with it, we will use an upper index to indicate time series related to a group: $X_i^l = \{x_i^l(1), x_i^l(2), \dots, x_i^l(T_i)\}$, $S^l = \{X_1^l, X_2^l, \dots, X_n^l\}$, $l \in \mathcal{N} = \{1, 2, \dots, N\}$. During extraction of subunits all elements in a group will be considered jointly, whereas different groups will be considered separately. For instance, one can assign one channel to one hand ($N = 2$) or one channel to one of the 14 features mentioned earlier ($N = 14$). In general, the number N as well as the assignment of the features to groups can be a subject of further research.

2.2 Sequence Partitioning Problem

Let us consider a time decomposition D^l , which, for each $i \in I$, defines a number $k_i^l = k_i^l(D^l) \geq 1$ and k_{i-1}^l cut points $t_{ij}^l = t_{ij}^l(D^l)$, where $1 < t_{i1}^l < t_{i2}^l < \dots < t_{i, k_i^l}^l < T_i$. The decomposition means that X_i^l is partitioned into k_i^l subsequences. The first subsequence $s_{i1}^l(D^l)$ starts at $t = 1$ and ends at $t = t_{i1}^l$, the next subsequence $s_{i2}^l(D^l)$ starts at $t = t_{i1}^l$ and ends at $t = t_{i2}^l$, and so on until the last subsequence $s_{i, k_i^l}^l(D^l)$ which starts at $t = t_{i, k_i^l}^l$ and ends at T_i . The resulting data set $S^{n^l}(D^l) = \{s_1^{n^l}(D^l), s_2^{n^l}(D^l), \dots, s_n^{n^l}(D^l)\}$, where $s_i^{n^l}(D^l) = \{s_{i1}^l(D^l), \dots, s_{i, k_i^l}^l(D^l)\}$, $i \in I$, contains $n^{n^l} = n^{n^l}(D^l) = \sum_{i=1}^n k_i^l(D^l)$ subsequences. The length of each subsequence is constrained by the minimal L_{min} and the maximal L_{max} number of points. We propose determining a good decomposition into subsequences by solving a multicriteria decision problem, based on the following main steps: (i) partition the set $S^{n^l}(D^l)$ into m^l (a given number) clusters, i.e. $S^{n^l}(D^l) = \{C_1^l(D^l), C_2^l(D^l), \dots, C_{m^l}^l(D^l)\}$, (ii) evaluation of the decomposition D^l using a vector of two criteria (indices) $J(D^l) = [J_1(D^l), J_2(D^l)]$ which characterizes the quality of the resulting clusters. The first criterion is the conditional entropy minimized by the minimum entropy clustering (MEC) algorithm described in [5]. Experiments presented in [5] show that MEC performs significantly better than k-means clustering, hierarchical clustering, SOM and EM. Moreover, it can correctly reveal the structure of data and effectively identify outliers simultaneously. To compare discrete sequences we use dynamic time warping (DTW) [6, 16]. DTW aligns two sequences while attempting to achieve the minimal difference. The warping path with the optimal distance d_{DTW} can be obtained by dynamic programming. The second criterion is the Dunn index DI [6, 7]. It is defined by two parameters: the diameter $diam(C_i^l)$ of the cluster C_i^l and the set distance $\delta(C_i^l, C_j^l)$ between C_i^l and C_j^l , where

$$diam C_i^l = \max_{x, y \in C_i^l} \{d(x, y)\}, \delta(C_i^l, C_j^l) = \min_{x \in C_i^l, y \in C_j^l} \{d(x, y)\} \quad (1)$$

and $d(x, y)$ indicates the distance between points x, y .

$$DI = \min_{1 \leq j \leq m^l} \left\{ \min_{1 \leq i \leq m^l, i \neq j} \left\{ \frac{\delta(C_p^l, C_q^l)}{\max_{1 \leq k \leq m^l} diam C_k^l} \right\} \right\} \quad (2)$$

Larger values of DI correspond to good grouping with compact and well separated clusters.

2.3 Optimization Method

As follows from subsection 2.2 our problem is a multiobjective optimization problem (MOP) with two criteria. To solve MOPs evolutionary algorithms are often used. Evolutionary algorithms deal simultaneously with a set of possible solutions (the so-called population) which allow us to find several members of the Pareto optimal set in single run of the algorithm [9]. Our approach to solve the MOP adopts the immune-based algorithm CLONALG originally used for ordinary optimization [9, 10]. We use lexicographic ordering [8]. Here the single objective J_1 (considered the most important) is optimized without considering J_2 . Then the J_2 is optimized but without decreasing the quality of the solution obtained for J_1 . In the sequel we shortly describe the algorithm, the encoding method, and the mutation operator.

CLONALG. The main loop (repeated gen times, where gen is the number of generations) consists of four main steps: one initial step where all the elements of the population are evaluated and three transformation steps: clonal selection, mutation, apoptosis.

1. Evaluation. For each element D^l in the population P compute $J_i(D^l)$, $i = 1, 2$ and perform lexicographic ordering of the elements.
2. Clonal selection. Choose a reference set $P_a \subset P$ consisting of h elements at the top of the ranking obtained in step 1.
3. Mutation.
 - a) For each $D^l \in P_a$ make c mutated clones Dc_j , $j = 1, 2, \dots, c$, compute their values $J_1(Dc_j)$, $J_2(Dc_j)$, and place the clones in the clonal pool CP .
 - b) Lexicographically order the elements of $P \cup CP$, choose a subset $P_c \subset P \cup CP$ containing B best elements, where B denotes the size of P .
4. Apoptosis. Replace b worst elements in P_c by randomly generated elements.
5. Set $P \subset P_c$.

In the algorithm the current population P is mixed with the clonal pool CP and the predefined number of best elements (i.e. at the top of the ranking) is picked up to form new population. The last step of the main loop replaces b worst solutions by randomly generated elements.

Encoding and mutation. Each element of the population P represents a decomposition D^l of the set S^l into a set S'^l (see section 2.2). It has the form of the integer valued vector $D^l = [t_{11}^l, t_{12}^l, \dots, t_{1,k_1^l-1}^l, t_{21}^l, t_{22}^l, \dots, t_{2,k_2^l-1}^l, \dots, t_{n1}^l, t_{n2}^l, \dots, t_{1,k_n^l-1}^l]$ composed of the cut points of the original sequences.

The mutation process consists of a given number M of mutations conducted on a population element. The mutation means an operation randomly chosen from the following variants: (a) add cut point (probability 1/4), (b) remove cut point (probability 1/4), (c) move cut point (probability 1/2). In

all cases a subsequence is randomly selected and, depending on a drawn variant, it is: (a) divided into two shorter subsequences, (b) joined together with its preceding subsequence, (c) made shorter or longer by shifting its initial point. New cut point in (a) and (c) is placed in a position randomly chosen from the corresponding set of feasible points, i.e. the points for which the resulting subsequences satisfy the length constraints. Similarly, the union in (b) is accepted if the resulting subsequence is not too long.

The optimization results in obtaining a good decomposition D_{opt}^l . We can use it to transform each sequence X_i^l to a string of labels $\tilde{X}_i^l = \{e_{i1}^l, e_{i2}^l, \dots, e_{i,k_i}^l\}$, where $e_{ik}^l \in E^l = \{\alpha_1^l, \alpha_2^l, \dots, \alpha_{m^l}^l\}$, α_k^l denotes the label assigned to the cluster C_k^l , and e_{ik}^l is a label of the cluster the subsequence $s_{ik}^l(D_{opt}^l)$ belongs to. Let us denote by \tilde{X}_i the string representation of X_i , i.e. $\tilde{X}_i = \{\tilde{X}_i^1, \tilde{X}_i^2, \dots, \tilde{X}_i^N\}$ and, consequently, by \tilde{S} the counterpart of S .

3 Subunit-Based Recognition

Let us assume that a word to be classified is represented by a sequence $Y = \{y(1), y(2), \dots, y(T_y)\}$. The feature vectors $y(\cdot)$ have the same structure as $x(\cdot)$ and therefore the sequences $Y^l = \{y^l(1), y^l(2), \dots, y^l(T_y)\}$, where $l \in \mathcal{N}$, will be considered separately. Two problems have to be solved. The first problem consists in finding an appropriate string representation of Y^l , i.e. $\tilde{Y}^l = \{e_{y1}^l, e_{y2}^l, \dots, e_{y,k_y^l}^l\}$, where $e_{ik}^l \in E^l$ and, consequently, the string representation \tilde{Y} of Y . The second problem is to find $NN(\tilde{Y})$ – the nearest neighbor of \tilde{Y} in the set \tilde{S} . Then the unknown word is assigned to the class which $NN(\tilde{Y})$ belongs to.

The string representation can be found by solving an optimization problem with respect to cut points of Y^l for each $l \in \mathcal{N}$. Let $D_y^l = [t_{y1}^l, t_{y2}^l, \dots, t_{y,k_y^l-1}^l]$ characterizes a decomposition. As opposed to the previous optimization, now the criterion to be minimized is $J(D_y^l) = \sum_{k=1}^{k_y^l} d_{DTW}(k)$, where $d_{DTW}(k)$ denotes the DTW distance between the k -th subsequence $s_{y,k}^l(D_y^l)$ of Y^l and its nearest neighbor $NN(s_{y,k}^l(D_y^l))$ in the set $S^l(D_{opt}^l)$. The optimization task can be solved by CLONALG. Then e_{jk}^l is a label of the cluster the $NN(s_{y,k}^l(D_{opt}^l))$ belongs to. The procedure is repeated for each $l \in \mathcal{N}$. The second problem is also an optimization task. Here the so called edit distance [6] is used as a measure of the difference between two strings. The method resembles DTW. It uses dynamic programming to find a minimum number of operations (insert, delete, replace) required to transform one string into the other. Let us denote by d_i^l the edit distance $d_{ED}(\tilde{Y}^l, \tilde{X}_i^l)$ between the string \tilde{Y}^l and a string \tilde{X}_i^l . The similarity measure between the sequence Y and a sequence X_i is the sum $d_i = \sum_{l=1}^N w^l d_i^l$, where w^l denotes a weight assigned to the l -th component of the feature vector. In particular, all the weights are

equal to one. The sequence Y becomes assigned to the class X_j belongs to, where $j = \arg \min_{i \in I} (d_i)$.

4 Experiments

In this section we present the results of experiments based on real sequences obtained for signed Polish words. The sequences represent 100 words which can be used at the doctor's and in the post office. Each word is characterized by a vector of 14 features mentioned in section 2.1. We used a data set consisting of sequences of feature vectors for 40 realizations of each of the 100 words. Gestures have been performed by two signers. One person is a PSL teacher, the other has learnt PSL for purposes of this research. Each signer repeated each word 20 times. The data have been registered with the rate 25 frames/s. To perform cross-validation we divided the data set into four disjoint subsets. Each subset consisted of data corresponding to 10 repetitions of each word (5 repetitions performed by each signer). We performed four experiments using three subsets as the training set S and the remaining subset as the test set. The experiments have been labeled with A, \dots, D . In each experiment the data in S were used to extract the subsequences as described in section 2 and the remaining elements were classified by the method described in section 3. Sample results of recognition are given in table 1. Subunits for each feature were extracted independently ($N = 14$), thus 14 symbolic transcriptions were assigned to each word in S . Parameters used by immune algorithm were as follows: $B = h = 10; c = 5; b = 2; M = 2; gen = 5; L_{min} = 4; L_{max} = 8$. We solved the optimization task for $m^l = 10$ clusters. Exemplary subsequences, obtained for the horizontal placement of the right hand center, extracted from words in the training set, with related symbolic transcriptions are shown in Fig. 1. Automatically determined subsequences' boundaries are marked with crosses. Resulting symbolic transcriptions based on ten subunits are given in brackets. Immune algorithm for creating symbolic transcriptions of the sequences from the test

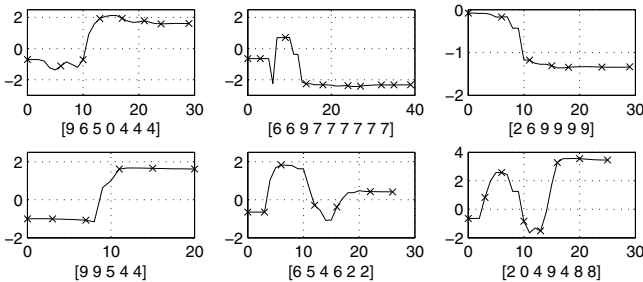


Fig. 1. Sequences representing six chosen words from a training set (from left to right, from top to bottom: *pharmacy*, *audiogram*, *angina*, *ambulance*, *parcel*, *man*)

Table 1. Results of the cross-validation test. Recognition rate in %. Because of randomness of the optimization algorithms each test was repeated ten times.

	A	B	C	D	Cross-validation mean
Minimum	82,0	92,4	91,8	89,4	
Maximum	85,7	94,2	95,2	92,1	
Mean	84,2	93,0	93,0	90,3	90,1
StDev	1,6	0,7	1,3	1,0	

set used identical parameters as those for extracting subunits. Table 1 shows the recognition rates. Results are quite promising and comparable with the results we obtained on the same dataset using heuristic definitions of visually-oriented subunits and related parallel hidden Markov models [18].

5 Conclusions

Large-vocabulary systems of sign language recognition require the modeling of signed expressions in smaller units than words. However, an additional knowledge of how to break down signs into subunits is needed. In vision-based systems the subunits are related to visual information. As linguistic knowledge about the useful partition of signs in regard of sign recognition is not available, the construction of an accordant partition is based on a data-driven process when signs are divided into segments that have no semantic meaning – then similar segments are grouped and labeled as a subunit.

In this paper we propose a new approach to determining the subunits. Subunits' boundaries are considered as decision variables in a multiobjective optimization problem. We use two objective functions, entropy and the Dunn index, as measures of cluster quality. These functions are optimized simultaneously. The method has been successfully verified using a database of 100 Polish words, but there remain some open questions concerning eg. the number of clusters, cluster validity indices, optimization methods etc. We will consider these issues in future research. A next step will be related to more advanced experimentation including recognition words and sentences of PSL. Interesting questions concern the choice of features and assignment of the features to groups during definition of subunits (section 2.1), as well as dependence of new words recognition on signer and size of training sets.

Acknowledgement

This research was supported by the Polish Ministry of Higher Education under grant N N516 369736.

References

1. Kraiss, K.F.: Advanced man-machine interaction. Springer, Berlin (2006)
2. Ong, S.C.W., Ranganath, S.: Automatic sign language analysis: a survey and the future beyond lexical meaning. *IEEE Trans. PAMI* 27, 873–891 (2005)
3. Vogler, C., Metaxas, D.A.: Framework for recognizing the simultaneous aspects of american sign language. *Computer Vision and Image Understanding* 81, 358–384 (2001)
4. Han, J., Awad, G., Sutherland, A.: Modelling and segmenting subunits for sign language recognition based on hand motion analysis. *Pattern Recognition Letters* 30, 623–633 (2009)
5. Li, H., Zhangm, K., Jiangm, T.: Minimum entropy clustering and applications to gene expression analysis. In: 3rd IEEE Computational Systems Bioinformatics Conference, pp. 142–151 (2004)
6. Xu, R., Wunsch, D.C.: Clustering. J. Wiley and Sons, Inc., Hoboken (2009)
7. Maulik, U., Bandyopadhyay, S.: Performance evaluation of some clustering algorithms and validity indices. *IEEE Trans. PAMI* 24, 1650–1654 (2002)
8. Miettinen, K.M.: Nonlinear multiobjective optimization. Kluwer Acad. Publ., Dordrecht (1998)
9. De Castro, L.N., Von Zuben, F.J.: Learning and optimization using the clonal selection principle. *IEEE Trans. on Evolutionary Computation* 6, 239–251 (2002)
10. Trojanowski, K., Wierzchon, S.: Immune-based algorithms for dynamic optimization. *Information Sciences* 179, 1495–1515 (2009)
11. Bicego, M., Murino, V., Figueiredo, M.A.T.: Similarity-based classification of sequences using hidden markov models. *Pattern Recognition* 37, 2281–2291 (2004)
12. Liao, T.W.: Clustering of time series data – a survey. *Pattern Recognition* 38, 1857–1874 (2005)
13. Handl, J., Knowles, J.: An evolutionary approach to multiobjective clustering. *IEEE Trans. on Evolutionary Computation* 11, 56–76 (2007)
14. Saha, S., Bandyopadhyay, S.: A symmetry-based multiobjective clustering technique for automatic evolution of clusters. *Pattern Recognition* 43, 738–751 (2010)
15. Tseng, V.S., Chen, C.H., Huang, P.C., Hong, T.P.: Cluster-based genetic segmentation of time series with DWT. *Pattern Recognition Letters* 30, 1190–1197 (2009)
16. Ratanamahatana, C.A., Keogh, E.: Three myths about dynamic time warping data mining. In: *SIAM Int. Conf. on Data Mining*, pp. 506–510 (2005)
17. Minimum Entropy Clustering Java package, <http://www.cs.ucr.edu/~hli/mec/>
18. Kapuscinski, T., Wysocki, M.: Recognition of signed Polish words using visually oriented subunits. In: *3rd Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, Poznan, pp. 202–206 (2007)

Extending 3D Shape Measurement with Reflectance Estimation

Robert Sitnik, Jakub Krzesłowski, and Grzegorz Mączkowski

Institute of Micromechanics and Photonics

Warsaw University of Technology

8 Boboli, Warsaw, PL 02-525

e-mail: r.sitnik@mchtr.pw.edu.pl, j.krzeslowski@mchtr.pw.edu.pl,

e-mail: g.maczkowski@mchtr.pw.edu.pl

Summary. The increasing need of providing information about the surface of complex 3D objects accounts for new approaches in shape, color and reflectivity measurements. A method integrating 3D shape and reflectance measurement has been developed, based on multispectral imaging and directional illumination. The unified data representation of objects under investigation can aid machine vision, digital documentation and expand means of realistic imaging of unique items in virtual reality. While there already exist methods of merging image based reflectance measurements with 3D data collected independently, this is a step toward integration and increase of data correspondence. Also, the proposed data processing method allows further extension, and fulfills needs of both an accurate and mobile solution.

1 Introduction

The evolution of image processing and analysis allows for endless applications, and as the age of information ascends, many approaches have been investigated and verified. This is especially true in digital world, if we take into account the exponential growth of computational power, optoelectronic device miniaturization and fluent knowledge exchange using the Internet. A parallel benefit is the ability to share results in the form of digital representation of objects in a compound state and with more details.

3D technology has been in use for some time now, both in engineering and digital reconstruction. Art conservators use it to create copies of precious artifacts, while in industry it is used for manufacturing process control [1], reverse engineering [2] and machine vision [3]. In medical applications, human body is a subject examined by the very similar techniques [4]. Image analysis and synthesis beyond 2D helps by allowing to estimate material and color which in turn allows to approximate the true nature of investigated surface. Existing measurement methods were developed independently on common grounds, leaving out the effort of combining results together in final stage of data processing. This however caused ambiguity in further analysis

and feature recognition, and raises questions on information correspondence. Combined efforts of mentioned representative groups prove, that there is a great urge for integration of modalities in shape measurement, as it is already showing in complexity of models used in computer graphics [5]. Such models can be created using single measurement setup described in this paper.

2 Integrated Measurement Method

The presented measurement method is based on structured light projection for shape measurement [6] extended with multispectral color measurement [7] and angular reflection measurement. These three modalities are realized in a single, integrated setup which is based on the use of CCD camera as the only needed input device. Furthermore, components of the setup are treated in modular fashion - can be easily switched off, exchanged, and are controlled by a single PC unit with a specially developed software, both for conducting calibration and measurements.

The calculation environment allows for processing of large data sets and storing them in specially designed structure [6]. Each cloud of points from shape measurement has assigned additional data layers which can hold information on reflectance, color or other properties. Automation of the process from data acquisition to surface model retrieval is done automatically with a set of parameters predefined by the user. The described data representation can be used to create a triangle mesh model of the 3D object in the same automated manner.

The whole system is calibrated in consecutive steps as shown in figure 1. First, the camera's geometrical calibration compensates detector distortion and correlates (i, j) pixel coordinates with (x, y, z) coordinates in 3D space. Then, the phase calibration is performed using the same reference plane object [8]. Third step is an intensity calibration which is necessary for color measurement stage because it provide information on emission spectrum of a light source used. Last step of the calibration process allows for a compensation of illumination non-uniformity in the field of view of the detector.

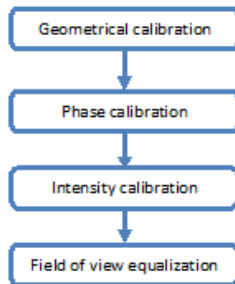


Fig. 1. Steps of a calibration procedure

Once the whole setup is calibrated, it becomes independent of the surrounding environment and can be reused for multiple measurements, as long as its configuration remains unvaried.

2.1 Shape Measurement

The shape measurement system relies on geometrical camera calibration which allows for compensation of its distortion and for correlation between camera’s pixels and scene’s coordinates [8]. Additionally a set of sine fringe patterns shifted in phase are projected on the measured surface in order to obtain relation between (x, y, z) coordinates in measurement volume and (i, j, Φ) coordinates registered by the camera. In result a cloud of points resembling shape of a measured surface is calculated.

2.2 Color Measurement

For determining the spectral reflectance properties of investigated surface, a light source of continuous spectrum is used together with a set of narrow band-pass filters. In the third calibration stage mentioned above, which uses a diffusive white reference plane with constant spectral reflectance in the field of view, the spectral characteristic of light source and the camera is compensated. As shown in equation (1), this process can be viewed as a linear product of 3 independent factors, where I_s is the desired variable.

$$I_d(\lambda) = I_i(\lambda)I_s(\lambda)I_a(\lambda), \tag{1}$$

where:

- $I_d(\lambda)$ – light intensity registered by the detector,
- $I_i(\lambda)$ – light intensity emitted by the light source,
- $I_s(\lambda)$ – spectral characteristic of the measured surface,
- $I_a(\lambda)$ – spectral sensitivity of the detector and optical system attenuation.

The non-uniformity of detector’s field of view due to vignetting and spectral filters angular characteristic diversity is compensated in the last step of the calibration process. A correction map of scaling coefficients is saved and applied to raw bitmap data after frame capturing. The spectral characteristic of every filter measured in advance is applied to create the map of energy distribution in the visible spectrum, so that variable filter transmittance can be accounted for. Then, for every pixel a profile is created from which a spectral reflectance is calculated according to equation (2):

$$I_s(i, j, \lambda) = I_d(i, j, \lambda) \frac{C}{I_{s \max}} \frac{1}{I_i(i, j, \lambda)I_a(i, j, \lambda)T(\lambda)}, \tag{2}$$

where:

C – constant corresponding to the white reference plates spectral characteristic,

$I_{s \max}$ – average maximum intensity from all spectral channels (normalization factor),

T – transmission extremes of spectral filters (spectrum sampling function) [%].

From estimated spectral reflectance an International Commission on Illumination (CIE) XYZ and $L^*a^*b^*$ color coordinates can be recovered [9].

2.3 Angular Reflection Measurement

An angular reflection measurement method relies on a set of evenly distributed light sources which can illuminate a surface under investigation from several different angles and thus images acquired this way can serve for scattering characteristic estimation.

Because the samples are averaged along the visible spectrum, the angular reflectance can be interpreted as the square of total reflectivity in visible spectrum. The spectral characteristic of used light sources is close to standard illuminant of natural daylight. The non-uniformities of their angular light distribution and unequal intensities are calibrated using the uniformly colored reference plane. Using the calculated normal vectors of surface points from shape measurement, as well as light source and camera geometry from calibration, the algorithm uses iterative nonlinear solver [10] to fit parameterized Bi-directional Reflectance Distribution Function (BRDF) [11] to data samples. It minimizes the resulting RMS error by modifying Phong coefficients k_d , k_s and e_s (equation [3]). The algorithm then saves best fitted coefficients for every pixel as additional data layer in a cloud of points. The overall error of fitting is also saved as a measure of uncertainty for every point. The parameterized reflectance distribution is given in the following equation:

$$\rho(\mathbf{i}, \mathbf{n}, \mathbf{v}, \mathbf{r}) = k_d \cdot \mathbf{i} \cdot \mathbf{n} + k_s \cdot (\mathbf{v} \cdot \mathbf{r})^{e_s} \quad (3)$$

where \mathbf{i} is the illumination direction, \mathbf{n} - the normal vector, \mathbf{v} - the viewing direction and \mathbf{r} is the direction of reflection.

Because of the time constraints for measurements outside of the laboratory, not the whole bidirectional reflectance distribution function (which would require scanning in five dimensions) is being measured. Instead, spectral and angular characteristics in every point on surface are measured separately. This assumes, that the measured surface has isotropic properties, that is, the glossiness does not vary with light wavelength.

The final integration of calculated reflectance parameters takes place in calculation of texture in the $L^*a^*b^*$ color space, where as lightness component L^* the angular reflectance characteristic is used and the a^* and b^* chromacity coordinates are calculated from spectral data [12]. The diagram in Fig. 2

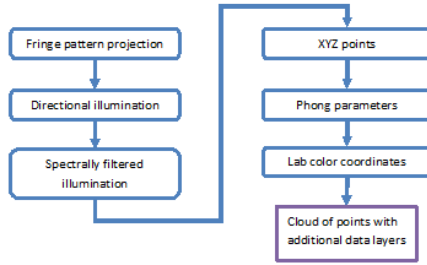


Fig. 2. Data acquisition and processing in consecutive steps

shows measurement stages and corresponding data which are acquired from them.

2.4 Merging Multiple Measurements

To create a full 3D object, several surface measurements from different directions are necessary. This can be automated by use of a rotation stage, while the developed algorithm allows fitting 3D data measured from multiple directions without the a priori knowledge of their relative orientation [8]. The algorithm uses curvature for finding initial transformation and Iterative Closest Point (ICP) [13] as the second step of precise fitting. Once the surface profiles are fitted together, the joint regions, common to more than one measurement direction can be merged using the local neighborhood of every point, as showed in Fig. 3. The quality coefficients are used for weighting the

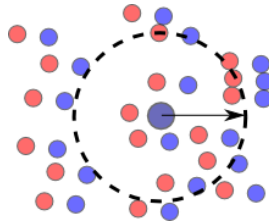


Fig. 3. Finding analyzed points’ neighborhood within a predefined radius. Cloud 1 - red, cloud 2 - blue.

mean value of calculated Phong parameters. The resulting data layer values are based on local average from several single direction measurements, for which the fitting error was smaller. This is shown in equation 4:

$$K(P) = \frac{\sum_{P_i: Q(P_i) > Q(P)} K(P_i) \cdot Q(P_i)}{\sum_{P_i} Q(P_i)} \quad (4)$$

where the calculated parameter K at an analyzed point P is changed to the mean value of surrounding points which represent a better estimation (higher Q values).

3 Experimental Setup

An exemplary setup configuration was used for performing test measurements. It comprised of a grayscale 1024×768 pixels Pointgrey Flea II CCD camera with 8mm lens and a digital light projector of the same pixel resolution. A 20nm spectral filter of 10nm bandwidth mounted on a circular filter changer in front of the camera lens and a Canon SpeedLite 580EX II flash were used for multispectral imaging. Eleven light emitting diode (LED) sources, distributed at even angles around the camera were used for directional reflectance measurement. The measurement scene was confined in a $400 \times 400 \times 600\text{mm}^3$ volume. A single directional measurement takes about 2 minutes, and is proportional to the sum of exposure time of collected frames. Lowest gain factor of the CCD camera was used for noise minimization. Fig. 4 shows an example appearance of the measurement system:

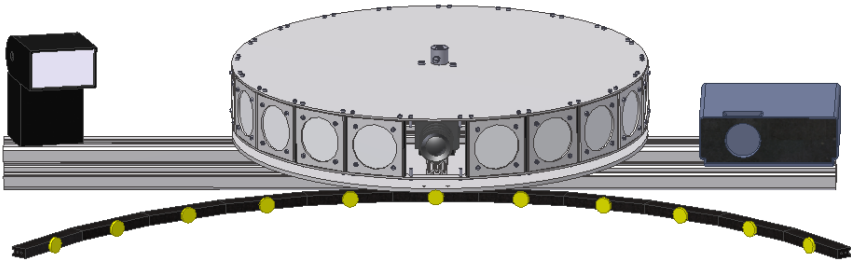


Fig. 4. The camera extension encompassing light source array, a flash and a rotary filter changer

4 Measurement Results

Measurement uncertainty of 3D shape measurement in developed system has been tested by the measurement of known calibration sample and it is equal to 0.04mm . To verify the angular measurement, a sample with diffusive surface was measured with a gonioreflectometric device. To both of the resulting BRDF slices a parameterized reflectance distribution function was fitted (Fig. 5). The overall error of Phong parameters estimation difference was within a 10% range. It should also be mentioned that surfaces of high glossiness, for which the angular resolution (resulting from light sources' distribution) is too sparse, are difficult to be measured by means of structured light

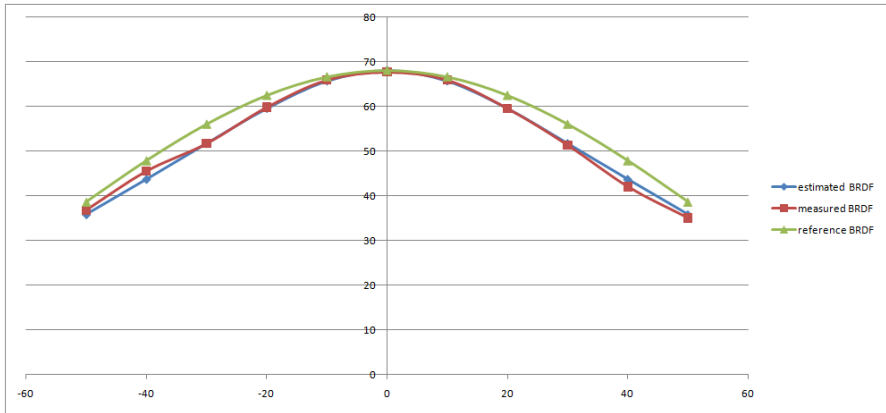


Fig. 5. Comparison of fitting a parameterized reflectance distribution to measurement and reference data of exemplary diffusive surface

projection and are beyond the scope of this work. To verify color measurement accuracy $L^*a^*b^*$ coordinates for every color patch of a Gretag Makhbeth ColorChecker target were calculated for the multispectral camera and a reference spectrophotometer, for D65 standard illuminant. Corresponding ΔC color differences [14] were also derived. The estimated standard deviation of color coordinates in $L^*a^*b^*$ color space is approximately $s = 5.28$. The mean color difference between color patches was calculated with the use of the ΔC , because the main goal was to estimate the accuracy of chromacity regardless of the level of signal which may differ in consecutive measurements. Sample results are summarized in Table 1. All reference values come from a measurement conducted with Minolta CM-2600d spectrophotometer.

Table 1. Example color differences obtained from GretagMakhbeth Color Checker evaluation

color patch	reference a^*	measured a^*	reference b^*	measured b^*	ΔC
light skin	16.82	13.75	16.71	18.92	3.78
blue sky	-1.38	-4.90	-21.08	-16.17	6.05
moderate red	45.87	43.49	15.75	15.30	2.42
green	-41.15	-36.73	30.67	28.39	4.98

Figure 6 shows a digital model (cloud of points) of a statuette of the ancient Greek goddess Kybele with surface appearance reconstructed with presented method.



Fig. 6. Example measurement result

5 Conclusions

An effort was made to develop a method of shape and reflectance measurement by integrating already proven approaches such as structured light projection technique, multispectral color measurement and angular reflection measurement. The portability of possible designs allows the usage in wide range of applications, from product verification, reverse engineering to aided computer graphics. The proposed measurement setup was a compromise between design complexity and measurement resolution. The described method allows the use of arbitrary matrix detector and projector as well as different number of light sources and spectral filters. If a higher level of accuracy is required, a system of both higher spectral and angular resolution can be realized by increasing the number of light sources and spectral channels, however at a higher time cost.

Acknowledgements

This work was performed under the grant No. PL0097 financed by the Norwegian Financial Mechanism and EOG Financial Mechanism (2004-2009).

References

1. Bieri, L.S., Jacot, J.: Three-dimensional vision using structured light applied to quality control in production line. In: Proc. SPIE, vol. 5457, p. 463 (2004)
2. Kubik, D.L., Greenwood, J.A.: Development of photogrammetric methods of stress analysis and quality control. Technical Sciences University of Warmia and Mazury in Olsztyn 6, 73–100 (2003)

3. Francken, Y., Cuyppers, T., Mertens, T., Gielis, J., Bekaert, P.: High quality mesostructure acquisition using specularities. *Computer Vision and Pattern Recognition*, 1–7 (2008)
4. Sitnik, R., Witkowski, M.: Locating and tracing of anatomical landmarks based on full-field 4d measurement of human body surface. *Journal of Biomedical Optics* 13 (2008)
5. Jensen, H.W., Marschner, S.R., Levoy, M., Hanrahan, P.: A practical model for subsurface light transport. In: *Proc. SIGGRAPH 2001*, pp. 511–518 (2001)
6. Sitnik, R., Kujawinska, M., Woznicki, J.: Digital fringe projection system for large-volume 360-deg shape measurement. *Opt. Eng.* 41, 443–449 (2002)
7. Tonsho, K., Akao, Y., Tsumura, N., Miyake, Y.: Development of Goniophotometric Imaging System for Recording Reflectance Spectra of 3D Objects. In: *Proc. SPIE Color Imaging: Device-Independent Color, Color Hardcopy, and Applications VII*, vol. 4663, pp. 370–378 (2002)
8. Sitnik, R.: New method of structure light measurement system calibration based on adaptive and effective evaluation of 3D-phase distribution. In: *Proc. SPIE*, vol. 5856, pp. 109–117 (2005)
9. Wyszecki, G., Stiles, W.S.: *Color Science: Concepts and Methods, Quantitative Data and Formulae*. John Wiley & Sons, Chichester (2000)
10. Lafortune, E., Foo, S., Torrance, K., Greenberg, D.: Non-linear approximation of reflectance functions. In: *SIGGRAPH 1997 Conference Proceedings, Annual Conference Series*, pp. 117–126 (1997)
11. Phong, B.T.: Illumination for computer generated pictures. *Communications of the ACM* 18, 311–317 (1975)
12. Malacara, D.: *Color Vision and Colorimetry: Theory and Applications*. SPIE Press, San Jose (2002)
13. Besl, P.J., McKay, N.D.: A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14, 239–256 (1992)
14. Lee, L.C.: *Color Imaging Science*. Cambridge University Press, Cambridge (2005)

Software Framework for Efficient Tensor Representation and Decompositions for Pattern Recognition in Computer Vision

Bogusław Cyganek

AGH University of Science and Technology Al. Mickiewicza 30,
30-059 Kraków, Poland
e-mail: cyganek@agh.edu.pl

Summary. In this paper we present a novel software framework for efficient representation and manipulations of tensors which aims in minimizing data copying. Tensors are stored in the matricized form with simultaneous abstraction superimposed on tensor indices thanks to the proxy design pattern. The proposed software pattern was then used in computation of the Higher- Order Singular Value Decomposition. Finally, the whole framework was tested in the problem of static gesture recognition.

1 Introduction

Tensors are mathematical objects which are characteristics of the special transformation rules for change of the coordinate systems [1]. In data mining and other data related applications they found broad application as the multidimensional arrays of data. They showed to be very useful in analysis of multifactor processes in which the data decomposition methods, such as Higher-Order Singular Value Decomposition (HOSVD) or best-rank, play important roles [3][7]. Although many platforms have been developed for efficient tensor representations they usually lack sufficient flexibility in using different data types or, usually being written in MatLab, they do not fit well into other programming platforms [2][3]. In this paper we address the problem of efficient tensor representation and manipulation in software realizations. Our main goals can be summarized as follows:

1. Flexibility in accessing tensors as multidimensional arrays and flat data representations at the same time, depending on a context and without necessity of additional copies.
2. Efficient software/hardware processing.
3. Flexibility in element type selection and specializations (e.g. tensors with boolean or fixed-point data types, etc.).

The proposed tensor representation was used in the algorithm for computation of the HOSVD which requires tensors being represented in each of

its flat modes. Then, the developed framework which allows computation of the HOSVD was applied to the problem of static gesture recognition. The experimental results showed its flexibility in this, as well as in many other pattern recognition tasks.

2 Architecture of the Software Framework

A key concept lying behind tensor representation in computer memory as well as in many decompositions is tensor flattening, called also tensor matricization, which refers to tensor represented and manipulated as matrices. Tensor flattening can be defined as follows [7]:

Given an P -th order tensor $\mathcal{T} \in \mathbb{R}^{N_1 \times N_2 \times \dots \times N_P}$ the k -mode vector (or a fiber) of \mathcal{T} is a vector obtained from the elements of \mathcal{T} by varying only one index n_k while keeping all other indices fixed. If from \mathcal{T} a following matrix

$$\mathbf{T}_{(k)} \in \mathbb{R}^{N_k \times (N_1 N_2 \dots N_{k-1} N_{k+1} \dots N_P)} \quad (1)$$

is formed, then columns of $\mathbf{T}_{(k)}$ are k -mode vectors of \mathcal{T} .

In other words, the k -mode representation of a tensor assumes taking the k -th index off the set of tensor indices which then becomes a row index of the flat matrix representing that tensor. Its column index is a product of all $P-1$ left indices. However, where an element of the tensor is stored in memory depends on an assumed permutation order of these left indices, which can be set in $(P-1)!$ ways. However, from these plethora only two found applications in mathematics – the so called forward and backward cycle modes [2]. For example, for a 2-mode flattening of a 4D tensor, i.e. $P=4$, we obtain the following orderings of the other indices 3-4-1, and 1-4-3, for the forward and backward cycle modes, respectively. An example of a four dimensional $320 \times 240 \times 4 \times 3$ (columns \times rows \times frames \times color-channels) tensor with the forward 1-mode flattening is presented in Fig. 1a and Fig. 1b, respectively. However, for obvious reasons only three dimensions are shown. That is the fourth dimension, representing indexing over the RGB color channels, is not extended. In other words, the color pixels are treated and displayed as unities. If all dimensions were extended, then the flattening in Fig. 1b would be



Fig. 1. An example of a 4D tensor representing a color video scene (a). Shown only three dimensions. Practical flattening scheme with the forward cyclic mode of the tensor indices (b).

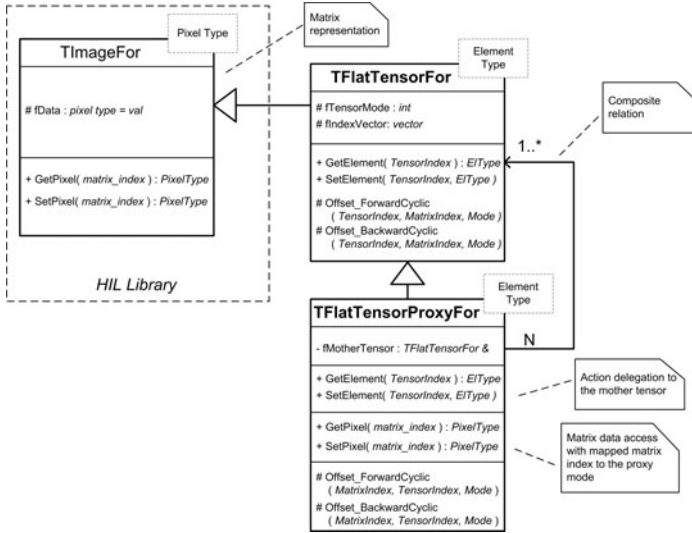


Fig. 2. A class hierarchy for storage and manipulation of tensors

tripled. The shown flattening with the forward cycle shows to be very practical in image processing since each frame of a video stream is kept together which allows easy conversions of the video streams into tensors.

Considering a square array of data, i.e. a matrix, the only one but substantial difference when compared with tensors is an external abstraction superimposed on a number and the way of permutations of tensor indices. This observation has a direct consequences in the presented software framework for tensor manipulations. Our main idea is to design a special class for manipulation of tensor indices instead of copying the whole array of data (which in practice can be huge) to obtain a different permutation cycle. This idea is obtained thanks to the proposed class hierarchy depicted in Fig. 2. The base template class *TImageFor*<> belongs to the HIL library, which is optimized for image processing and computer vision tasks but also for matrix operations [4][6]. It was also designed for as much as possible seamless hardware acceleration [5]. The main class for tensor representation is *TFlatTensorFor*<>, derived from *TImageFor*<> which serves to represent 2D matrices. Thus, in our framework a tensor is a specialized version of a matrix class, although this does not follow the usual mathematical way, in which a matrix is seen as a special two-dimensional tensor. However, our reversal is justified by the fact that a tensor in our system is always stored in the flattened representation with a chosen mode. Such a twofold approach expresses a computer storage capabilities (a computer memory is linear). In these terms a tensor is just a data structure (a matrix) with a tensor specific coordinate transformation scheme superimposed by the more specialized objects in the hierarchy. At the same time, however, we retain the

flexible and efficient methods developed for manipulation of the matrix data objects [6].

Specifically we'll be able to seamlessly manipulate data by hardware implementations. For example we can take advantage of the fast hardware SVD decomposition optimized for the objects of *TImageFor<>*. Thus, the "is a" relationship showed up to be a more practical solution than the previously tried "has a", in which a tensor was in possession of a matrix that stored its data in one mode. Thanks to the above, the *TFlatTensorFor<>* has two sets of the most important methods for accessing its elements. The first pair *Get/SetElement* takes as an argument a *TensorIndex* which is a vector of indices with length equal to the dimension of the tensor. The second set comprises the *Get/SetPixel* methods inherited from its base *TImageFor<>*, which allow access to the matrix data provided just row and column (*r,c*) indices.

The *TFlatTensorProxyFor<>* implements the proxy design pattern [4]. Such proxies are useful in all algorithms which require tensor representations in different flat *n*-modes. Proxy allows this *without* a necessity of creating a copy of the tensor which operation could easily consume significant portion of memory and time. An example of such an algorithm is the HOSVD decomposition, which is discussed in next section. In each its step a *n*-mode flat tensor needs to be obtained from the initial tensor T, for all possible *n*'s [7]. All these two-way index transformations are possible with the *Offset_ForwardCyclic / Offset_BackwardCyclic* methods which are able to recompute tensor-matrix indices in two ways and in two cyclic modes (backward and forward), as well as for different *n*-modes.

An index of an element of a tensor T of dimension *k* is given by a tuple (i_1, i_2, \dots, i_k) of *k* indices. This transforms into an offset *q* of a linear data structure as follows

$$q = (((i_1 n_2 + i_2) n_3) + \dots) n_k + i_k \quad (2)$$

where a tuple (n_1, n_2, \dots, n_k) defines dimensions of T. Naturally, matrix representation needs two selected dimensions $(r, c) = \left(n_m, \prod_{z=1, z \neq m}^k n_z \right)$, where *m* denotes a mode of the tensor. An element at index *q* needs to fit into such a matrix. In the tensor proxy object the problem is inverted, i.e. given a matrix offset *q* a corresponding tensor index tuple needs to be determined due to different modes of the tensors. This is done by successive division of the *q* in (2) by n_p for starting from $p = k$ up to $p=1$ (in C++ implementation indexing starts from 0), since it holds that for all *k*: $i_k < n_k$. A series of indices i_p is obtained in a form of residua of such successive divisions.

Summarizing, the advantages of the presented software component are as follows:

1. A uniform treatment of the tensor and its matrix *n*-mode representation. Thanks to this a tensor is represented only in a single chosen mode, and

it is not necessary to unfolding a tensor into its other flat representation and vice versa.

2. Application of the tensor proxy objects allows simultaneous manipulation of a tensor in its all possible n-mode flat representations without copying data.
3. The tensor component is embedded into the HIL library which allows hardware acceleration.
4. Template implementation allows different types of tensor elements. These can be further specialized for special types such as boolean or fixed-point data formats.
5. Object oriented C++ implementation can be easily ported to other object oriented languages such as C#, Java, Python, etc.

3 Computation of the HOSVD Tensor Decomposition

In this section we show the flexibility of our proxy representation when applied to one of the tensor decompositions. Analogously to the SVD [9] singular value decomposition (SVD) decomposition of the 2D tensors (matrices), there exists a P -th order decomposition of the P dimensional tensor \mathbf{T} . This is called the Higher-Order SVD singular value decomposition (SVD) (HOSVD) Higher order singular value decomposition decomposition of the P -th order tensors. It allows each tensor $\mathcal{T} \in \mathfrak{R}^{N_1 \times N_2 \times \dots \times N_m \times \dots \times N_n \times \dots \times N_P}$ to be decomposed as follows

$$\mathcal{T} = \mathcal{Z} \times_1 \mathbf{S}_1 \times_2 \mathbf{S}_2 \dots \times_P \mathbf{S}_P \quad (3)$$

where \mathbf{S}_k denotes a *mode matrix*, which is a unitary matrix of dimensions $N_k \times N_k$ spanning the column space of the matrix \mathbf{T} **matrix!translation** $^{(k)}$ obtained from the *mode-n* flattening of \mathcal{T} ; $\mathcal{Z} \in \mathfrak{R}^{N_1 \times N_2 \times \dots \times N_m \times \dots \times N_n \times \dots \times N_P}$ is a *core tensor* of the same dimensions as \mathcal{T} . Their properties are described in [7][8].

Based on the decomposition (3) the following sum can be constructed

$$\mathcal{T} = \sum_{h=1}^{N_P} \mathcal{T}_h \times_P \mathbf{s}_P^h \quad (4)$$

where

$$\mathcal{T}_h = \mathcal{Z} \times_1 \mathbf{S}_1 \times_2 \mathbf{S}_2 \dots \times_{P-1} \mathbf{S}_{P-1} \quad (5)$$

denote the basis tensors and \mathbf{s}_P^h are columns of the unitary matrix \mathbf{S}_P . Since \mathcal{T}_h is of dimension $P-1$ then \times_P in (4) is an outer product, i.e. a product of two tensors of dimensions $P-1$ and 1. The result is a tensor of dimension P , i.e. the same as of \mathbf{T} . In this case \mathcal{T}_h becomes two-dimensional, i.e. it is a matrix. Moreover, it is worth noting that due to orthogonality of the core tensor **tensor!core** \mathcal{Z} gn (5), \mathcal{T}_h are also orthogonal. Hence, \mathcal{T}_h in decomposition (4) constitute a basis. Thus, pattern recognition with HOSVD boils down to testing a distance of a given test pattern P_x to its projections in each of

the spaces spanned by the set of the bases \mathbf{T}_h in (4). This can be done by computing the following minimization problem

$$\min_{i, c_h^i} \left\| P_x - \sum_{h=1}^N c_h^i \mathbf{T}_h^i \right\|^2 \quad (6)$$

where c_h^i are the coordinates of P_x in the manifold spanned by \mathbf{T}_h^i , $N \leq N_P$ denotes a number of chosen dominating components. Due to the orthogonality of the tensors \mathbf{T}_h^i , the above reduces to the maximization of the following parameter [10]

$$\rho_i = \sum_{h=1}^H \left\langle \hat{\mathbf{T}}_h^i, \hat{P}_x \right\rangle^2 \quad (7)$$

where $\langle \cdot, \cdot \rangle$ is a scalar product of the tensors, $\hat{\mathbf{T}}_h^i$ and \hat{P}_x denote normalized components. In other words, returned is a class i for which the corresponding ρ_i from (7) is the largest. However, in our system we set a threshold ($\tau=0.85$) below which the system answers “don’t know”. Such a situation arises e.g. if a pattern is provided which the system was not trained for. The number N in (7) of components was set from 3 to 9. The higher N , the better fit, though at an expense of computation time. In our framework the original tensor \mathbf{T}_i of a class i is obtained from the available exemplars of the prototype patterns for that class (which can be of different number). These are patterns cropped from the training gesture images which are additionally rotated in a given range (in our experiments this was $\pm 12^\circ$ with a step of 2°) with additionally added normal noise. Such a strategy allows each pattern to be trained with different number of prototypes. Finally, the training stage ends in computation of \mathbf{T}_h^i for each \mathbf{T}_i , in accordance with (5).

The HOSVD algorithm relies on successive application of the matrix SVD decomposition [9] applied to each of the flattened $\mathbf{T}_{(k)}$ versions of the original tensor \mathcal{T} [7]. In result the \mathbf{S}_k matrices required in (3) are obtained. However, thanks to the proposed software framework, only one tensor with data is created whereas all flattened $\mathbf{T}_{(k)}$ are represented with the *TFlatTensorProxyFor* objects which do not require data copying.

4 Experiments

The proposed software framework for tensor representation, which for instance allows efficient computation of the HOSVD decomposition, was tested in static hand gestures recognition. Our data base of prototype gestures is presented in Fig. 3. It contains 12 predefined patterns which are then used to train the HOSVD classifier.

However, to cope with the possible deformations of the patterns expected in real life, the HOSVD classifier is trained with affinely deformed versions of the prototypes from the data base in Fig. 3. In our case we assumed possible



Fig. 3. The data base of prototype exemplars from which the rotated and noisy patterns were created, then used to construct the prototype tensors

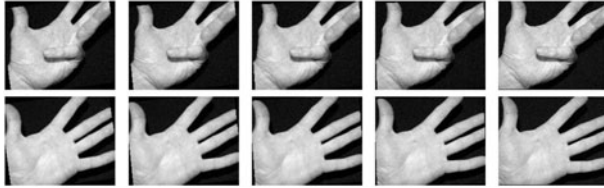


Fig. 4. Deformed versions of the original two prototype patterns from the gesture data base

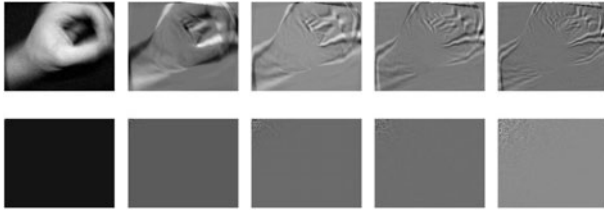


Fig. 5. First five tensors \mathcal{T}_h for a gesture (upper), and the corresponding five core tensors \mathcal{Z}_n with energy concentration spreading out from the left top corner (lower)

rotations with simultaneous small change of scale. Additionally, each transformed pattern, which two exemplars are shown in Fig. 4, is endowed with Gaussian noise which operation showed to improve accuracy of the system.

Fig. 5 and Fig. 6 show the first $N=5$ tensors \mathcal{T}_h for two different gestures (upper rows), and the corresponding core tensors \mathcal{Z}_n . It is interesting to observe that indeed, the energy is concentrated first in the top left corner, and then spreading out from this point.

The system is not able to identify all deformations, but for static gestures the error rate is in order of 15-25% which is acceptable for passing simple commands in our test applications for visual communication with a computer.

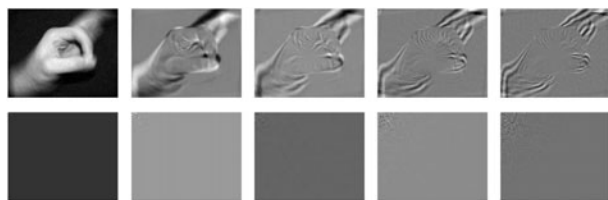


Fig. 6. First five tensors \mathcal{T}_h for another hand gesture (upper row), and the corresponding five core tensors \mathcal{Z}_n (lower row)

5 Conclusions

In this paper we present a novel software representation of tensors which allows their flexible representation and manipulation of their elements with no necessity of data copying. Thanks to this all algorithms which treat tensors in their flatten (matricized) representation also avoid data copying. An example is the HOSVD decomposition which was applied to the static gesture recognition problem with satisfactory results.

Acknowledgement

This work was supported from the Polish funds for scientific research in 2010.

References

1. Aja-Fernández, S., de Luis García, R., Tao, D., Li, X.: *Tensors in Image Processing and Computer Vision*. Springer, Heidelberg (2009)
2. Bader, B.W., Kolda, T.G.: MATLAB Tensor Classes for Fast Algorithm Prototyping. *ACM Transactions on Mathematical Software* 32(4), 635–653 (2006)
3. Cichocki, A., Zdunek, R., Phan, A.H., Amari, S.-I.: *Nonnegative Matrix and Tensor Factorizations*. Wiley, Chichester (2009)
4. Cyganek, B., Siebert, J.P.: *An Introduction to 3D Computer Vision Techniques and Algorithms*. Wiley, Chichester (2009)
5. Cyganek, B.: Architecture of an Integrated Software-Hardware System for Accelerated Image Processing. LNCS, vol. 5337, pp. 1–13. Springer, Heidelberg (2009)
6. http://www.wiley.com/legacy/wileychi/cyganek3dcomputer/supp/HIL_Manual_01.pdf
7. Lathauwer de, L.: *Signal Processing Based on Multilinear Algebra*. PhD dissertation, Katholieke Universiteit Leuven (1997)
8. Lathauwer de, L., Moor de, B., Vandewalle, J.: A Multilinear Singular Value Decomposition. *SIAM Journal Matrix Analysis and Applications* 21(4), 1253–1278 (2000)
9. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: *Numerical Recipes. The Art of Scientific Computing*. Cambridge University Press, Cambridge (2007)
10. Savas, B., Eldén, L.: Handwritten digit classification using higher order singular value decomposition. *Pattern Recognition* 40, 993–1003 (2007)

Hand Shape Recognition in Real Images Using Hierarchical Temporal Memory Trained on Synthetic Data

Tomasz Kapuściński

Rzeszow University of Technology,
Department of Computer and Control Engineering,
W. Pola 2, 35-959 Rzeszów, Poland
e-mail: tomekkap@prz-rzeszow.pl

Summary. In this paper the generalization capabilities of the Hierarchical Temporal Memory (HTM), the new computing paradigm based on cortical theory, has been exploited in order to recognize the hand shape in real images while training is done on synthetically generated data. It has been shown that the HTM trained in this way and tuned up with a small number of real examples gives pretty good recognition rates. Additionally the good scalability of the proposed solution has been observed while analyzing the recognition rates for the class that is 'almost unknown' because only few examples are shown during training.

1 Introduction

Humans can correctly recognize hand gestures despite changes in location, scale, lighting and in the presence of deformations and noise. Our brains have gain this ability by being exposed to some number of patterns corresponding to the particular gesture seen from different views. However, we can also properly classify the gesture even if the pattern that corresponds to it is not identical or close to these already seen.

Currently, despite numerous attempts to incorporate the same ability to computer systems (e.g. [1, 4, 8, 9, 10, 11]), the invariant gesture recognition still seems to be unsolved. In the most cases the problem is tackled by having training images, containing the gestures observed at different locations, distances, etc. This requires gathering the large data sets using procedures that are often inconvenient for the user and time consuming. Therefore the possibility of training process automatization using the synthetically generated images is particularly attractive.

For the practical reasons we are not able to generate images for the infinite number of possible transformations. Therefore the recognition tool having good generalization capabilities and scalability is required. Recently the Hierarchical Temporal Memory (HTM), the new computing paradigm, has been developed [5, 6]. It is based on the neuroanatomical research of the

mammalian neocortex and observations how the living creatures solve the vision problem [2, 3]. We can pretty well recognize objects from a single static snapshot but when we learn we use time extensively. As children, when we are confronted with a new and confusing object, we pick it up and move it about in front of our eyes. As the object moves, the patterns on our retina change and our cortical system is able to build the invariant model.

The invariant information about the object is distributed across many nodes up and down the hierarchy. Low-level visual details are stored in low-level nodes, and high-level structures are stored in higher-level nodes. Models created in our brain are composed from the smaller reusable units. This leads to good scalability and gives mechanism to solve ambiguities. As the information propagates up the hierarchy, it becomes more stable and unambiguous. In [7] HTM has been used to recognize synthetically generated gestures under large variation in hand's rotation.

This paper is organized in 5 sections. Section 2 contains a brief overview of HTM. Section 3 defines the problem and gives details of the proposed approach. Experiments, and obtained results are summarized and discussed in section 4. Section 5 concludes the paper with a summary and possible further works.

2 Hierarchical Temporal Memory Concept

Hierarchical Temporal Memory is a technology that replicates the structural and algorithmic properties of the neocortex. HTM is organized as a tree-shaped hierarchy of nodes. HTM receives the spatio-temporal pattern coming from the senses. Through a learning process it discovers what the causes are and develops internal representations of the causes in the world (Fig. 1). After an HTM has learned what the causes in its world are and how to represent them, it can perform inference. Given a novel sensory input, it infers what known causes are likely to be present in the world at that moment.

Each node in HTM implements a common learning and memory function. The basic operation of each node is divided into two steps (see Fig. 2). The first step is to assign the node input pattern to one of a set of quantization

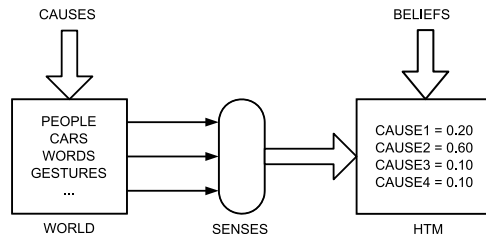


Fig. 1. HTM concept

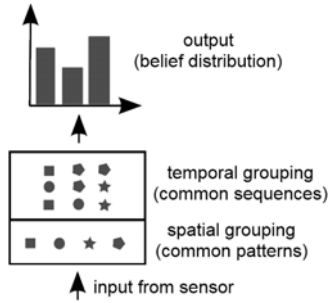


Fig. 2. HTM node operation: learning the invariant representations

points (spatial grouping). The node decides how close (spatially) the current input is to each of its quantization points. In the second step, the node looks for common sequences of these quantization points (temporal grouping). Discovered causes are connected to the common sequences stored in the node's temporal sublayer. These common sequences contains all learned variations of the observed object. Thanks to that mechanism different visual patterns belonging to the same object can be categorized as the same class because they occur close enough in time.

Given the static input we can obtain the distribution over all causes (in the noiseless case it is based on the occurrence of the given static pattern in the particular stored sequence). The detailed description can be found in [2].

3 Problem Definition and Proposed Solution

The 8 static hand gestures have been considered (Fig. 3(a)-(h)). Each gesture has been registered 200 times at varying hand position in 320 x 240 image frame. There are also some slight changes in the hand rotation in the obtained images, which are done unintentionally if the gesture is presented in a natural way.

For the training purposes the large database of the synthetic images, containing the corresponding gestures has been generated (Fig. 3(i)-(p)). Synthetic images have been created using the graphical tool Virtual Hand Studio, which gives pretty realistic views of the hand [12]. Each gesture has been shown at 4 different realizations, 36 hand rotations around its vertical axis (from 0 to 350 degrees with the regular step of 10) and 10 rotations around its horizontal axis (for 0 to 90 with the step of 10). Then, for the frontal view (both rotation angles equal to 0) of each gesture the affine transform, composed from translation and scale change, has been applied. Due to this transform each gesture was shown moving across the screen at three different scales. The whole training data set consisted of 691200 images, 86400 for each gesture.

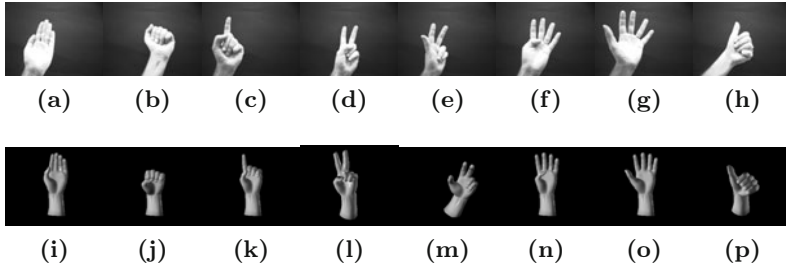


Fig. 3. Real (a - h) and synthetic (i - p) images of the hand gestures used in experiments

After many experiments with the network structure and parameters the following two-layer HTM has been proposed (Fig. 4).

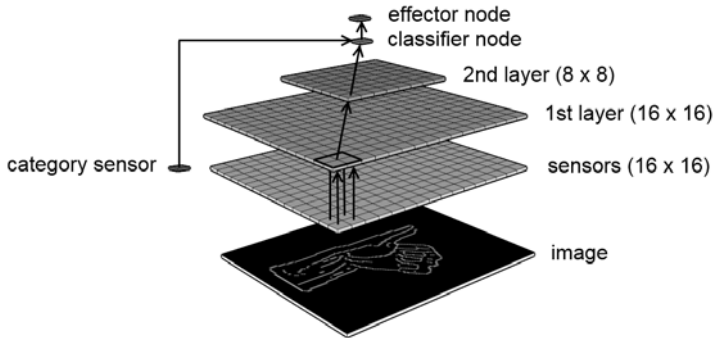


Fig. 4. The HTM topology

The sensors' layer consists of 16×16 elements. The edge orientation histogram, calculated from the 20×15 pixels image window covered by the sensor, has been chosen as the sensor's output. Only four different edge orientations are distinguished: horizontal, vertical and two slanting. The category sensor is connected only during the learning stage and it supplies the information about the gesture to which the current sensors reading belongs. The hierarchy is mapped by putting two layers containing 16×16 and 8×8 nodes respectively. Each of these layers consists of two sub-layers making spatial and temporary grouping appropriately [5, 6]. Above these layers the NN classifier and the effector node is put. The effector node writes the results to the file. The hierarchy manifests itself in linking the nodes between the particular layers.

4 Results and Discussion

The following three experiments have been carried out:

1. The HTM has been trained on the synthetically generated data (691200 images, 86400 for each gesture) and tested using the real data (1600 images, 200 for each gesture). The 1st row in the Table 1 shows the recognition result obtained for the testing set.
2. The memory trained in experiment 1, has been tuned up with a small number of randomly selected real data (64 images, 8 for each gesture). The 2nd row in the Table 1 shows the recognition results for the testing set composed from the remaining gestures (1536 images, 192 for each gesture).
3. The recognition of the gesture falling in category that is 'almost unknown' for the recognizer has been tested. The gesture *thumb* has been selected (Fig. 3h). The memory has been trained on the synthetically generated data with the exception of images belonging to the mentioned category. Then a small number of real examples of unknown gesture have been shown (8 randomly selected examples) and the testing has been carried out on the real images of the given gesture only (192 images).

Table 1. Recognition rates

experiment	recognition rate [%]
1	89.0
2	97.0
3	98.0

The experiments indicate that HTM trained on synthetically generated data can be used to recognize gestures in real images. It is also reasonable to consider the tuning up the network with small number of real examples. It is easy to achieve in practice because all the nodes used to construct the memory can work in training and inference mode simultaneously. Adding the new category to the system trained on other gestures is easy and can be achieved by showing only few examples of the new gesture (see Fig. 5). Experiment 3 was repeated for remaining gestures and the comparable results were obtained. The capabilities of the HTM noticed experimentally can be explained by analyzing the algorithm implemented in the nodes.

Consider a simple two-layer network with two nodes in the layer 1 and one node in the layer 2, as shown in Fig. 6(a).

This network is trained layer by layer. In the Fig. 6 and in the further explanations, the first superscript denotes the layer number, the second - the node number within this layer. The subscript denotes the spatial cluster

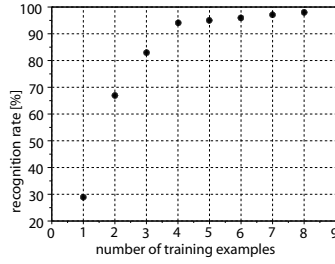


Fig. 5. Dependence of the recognition rate of *thumb* gesture on amounts of data used for training (HTM was first trained on the gestures belonging to the remaining categories)

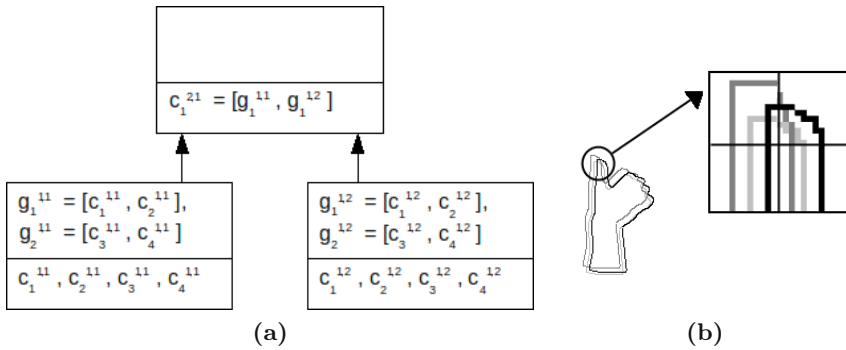


Fig. 6. Generalization mechanism

(or temporal group) number within the node. Suppose that the first layer is already trained. Each node in this layer has learned four spatial clusters (also called quantization centers): $c_1^{11}, c_2^{11}, c_3^{11}, c_4^{11}$ for the first node and $c_1^{12}, c_2^{12}, c_3^{12}, c_4^{12}$ for the second one. These spatial clusters have been grouped into two temporal sequences: $g_1^{11} = [c_1^{11}, c_2^{11}]$, $g_2^{11} = [c_3^{11}, c_4^{11}]$ for the first node and $g_1^{12} = [c_1^{12}, c_2^{12}]$, $g_2^{12} = [c_3^{12}, c_4^{12}]$ for the second one. Now, the first layer is switched to the inference mode and the training of the second layer starts. Suppose that the network sees the reading from sensors identified by the first layer as the quantization centers c_1^{11} and c_2^{12} . These centers will activate the temporal groups g_1^{11} and g_1^{12} respectively. The second layer node learns that this particular co-occurrence of the temporal groups is characteristic for the category being shown to sensors. This mechanism assures that all the patterns that correspond to the cross-product space of g_1^{11} from the left child and g_1^{12} from the right child will belong to the same category. All these input patterns will co-activate g_1^{11} from the left child and g_1^{12} from the right child.

It is remarkable that to learn this co-occurrence, the layer 2 had to see only one example.

The described mechanism has the following consequences. Consider Fig. 6(b). Assume that, the nodes in the layer 1 have all learned different transformations of vertical lines, upper-left corners and arcs as their temporal groups. The second layer memorized a thumb as a particular configuration of these temporal groups, by getting exposed to one image only (shown using the black color). The network is able to generalize to different scale variations and translations of this thumb.

HTM rapidly resolves conflicting or ambiguous input as information flows up the hierarchy. Imagine a network with three nodes, a parent node and two children nodes (Fig. 7).

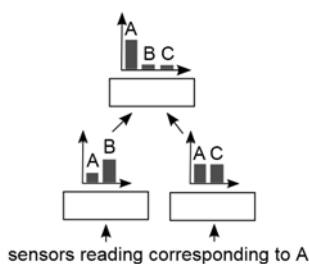


Fig. 7. Beliefs propagation in hierarchy helps to solve ambiguities

The first child node believes that it is seeing rather B, which is in fact wrong response. The sibling node's answer is ambiguous, it is not certain if it sees A or C. Parent node decides with high certainty that A is present. It chooses A because this belief is the only one that is consistent with its inputs. It made this choice even though A was not the most likely beliefs of the child nodes. Belief propagation in the hierarchy assures that the system very quickly settle.

5 Conclusions and Future Work

Hierarchical Temporal Memory trained on the synthetically generated data has been used for recognition of hand shape in real images. This tool has been chosen because it replicates the structural and algorithmic properties of the human neocortex, therefore it should be able to recognize the gestures in a way humans do. During the training HTM builds the invariant object's representations. At the inference phase it deals pretty well with ambiguities. The hierarchical structure of the HTM allows for sharing the representations. Different objects in the higher level can be composed from the same lower level parts. This leads to good scalability and storage efficiency. Further

works will include dealing with complex, non-uniform backgrounds. Because the nodes in the layer 2 respond to particular co-occurrences of the low level features the HTM should filter out undesirable details of the background. Some interesting additional human capabilities may also be explored, e.g. the mechanism of covering attention. Human system is able to limit the perceptual experience to a variable size area in the center of the visual field.

Acknowledgment

This research was partially supported by the Polish Ministry of Higher Education under grant N N516 369736.

References

1. Athitsos, V., Sclaroff, S.: Estimating 3D Hand Pose from a Cluttered Image. In: Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003), vol. 2, pp. 432–440 (2003)
2. Dileep, G., Bobby, J.: The HTM Learning Algorithms. Numenta Inc. (2007), <http://www.numenta.com/for-developers/education/algorithms.php>
3. Dileep, G., Hawkins, J.: Towards a Mathematical Theory of Cortical Microcircuits. PLoS Computational Biology 5(10) (2009)
4. Guan, H., Chang, J.S., Chen, L., Feris, R.S., Turk, M.: Multi-view appearance-based 3d hand pose estimation. In: Proc. of Conference on Computer Vision and Pattern Recognition Workshop, pp. 154–155. IEEE Computer Society, Washington (2006)
5. Hawkins, J., Blakeslee, S.: On Intelligence. Times Books, New York (2004)
6. Hawkins, J., Dileep, G.: Hierarchical Temporal Memory, Concepts, Theory, and Terminology. Numenta Inc. (2006), http://www.numenta.com/Numenta_HTM_Concepts.pdf
7. Kapuscinski, T.: Using Hierarchical Temporal Memory for Vision-Based Hand Shape Recognition under Large Variations in Hand's Rotation. In: Artificial Intelligence and Soft Computing part II. LNCS (LNAI), pp. 272–279. Springer, Heidelberg (2010)
8. Lee, S.U., Cohen, I.: 3D Hand Reconstruction from a Monocular View. In: 17th International Conference on Pattern Recognition (ICPR 2004), vol. 3, pp. 310–313 (2004)
9. Ong, E., Bowden, R.: A boosted classifier tree for hand shape detection. In: Proc. of 6th International Conference on Automatic Face and Gesture Recognition, pp. 889–894 (2004)
10. Stenger, B., Thayananthan, A., Torr, P., Cipolla, R.: Hand pose estimation using hierarchical detection. LNCS, pp. 105–116. Springer, Heidelberg (2004)
11. Tanimoto, T., Hoshino, K.: Real Time Posture Estimation of Human Hand for Robot Hand Interface. In: Second International Symposium on Universal Communication, pp. 303–308. IEEE Computer Society, Los Alamitos (2008)
12. Virtual Hand Studio. 3D Virtual Figure Drawing, Anatomy Softwares and Anatomical models. D-AnatomyStore, CloudStars, <http://d-anatomystore.com/>

Performance Comparison among Complex Wavelet Transforms Based Face Recognition Systems

Alaa Eleyan¹ and Hasan Demirel²

¹ Electrical & Electrical Engineering Department,
European University of Lefke, Turkey
e-mail: aeleyan@eul.edu.tr

² Electrical & Electrical Engineering Department,
Eastern Mediterranean University, Turkey
e-mail: hasan.demirel@emu.edu.tr

Summary. In this paper we investigate the recently developed dual tree complex wavelet transform (DT-CWT) and the single tree wavelet transform (ST-CWT) and compared them with Gabor wavelet transform for the face recognition problem. Experiments are carried out on standard databases. The resulting feature vectors of complex wavelets were applied to PCA and LDA for dimensionality reduction. In all experiments, complex wavelets equaled or surpassed the performance of Gabor wavelets in recognition rate when equal number of orientations and scales are used. Moreover, generally ST-CWT results outperformed DT-CWT. Obtained results indicate that complex wavelets can provide a successful alternative to Gabor wavelets for face recognition both using PCA and LDA.

1 Introduction

Identifying a person using geometric or statistical features derived from a face image is an important and challenging task [1, 2, 3, 4, 5]. This task becomes even more challenging due to the fact that large variations in the visual stimulus arising from illumination condition, viewing directions, poses, facial expression, aging, disguises are all common in real applications. A face recognition system should, to a large extent, take into account all the above mentioned natural constraints and cope with them in an effective manner. In order to achieve this, one must have efficient and effective representations for faces. Furthermore, it is also desirable that the representation derives its roots in some form from the principles of human visual processing. Gabor wavelet based representation provides an excellent solution and have been extensively studied in many image processing applications [6, 7, 8].

Lades *et al.* [9] used a dynamic link architecture framework of the Gabor wavelet for face recognition. Wiskott *et al.* [10] subsequently developed a Gabor wavelet-based elastic bunch graph matching (EBGM) method to label and recognize human faces. Liu *et al.* [11] proposed a method to determine the

optimal position for extracting the Gabor feature such that number of feature points are minimized while the representation capability is maximized. Liu and Wechsler [12] presented an independent Gabor features (IGFs) method.

Even though Gabor wavelet based face image representation is optimal in many respects, it has got two important drawbacks that shadow its success. First it is computationally very complex. Second, memory requirements for storing Gabor features are very high.

Complex approximately analytic wavelets provide a multiscale representation of images with good directional selectivity, invariance to shifts and in-plane rotation, and phase information much like the Gabor wavelets. The complex wavelets however are orthogonal and can be implemented with short one-dimensional separable filters which make them computationally very attractive. Unlike the Gabor wavelets, where the redundancy is 40 times with 5 scales and 8 directions, complex wavelet representation is 4 times redundant in 2 dimensions and the redundancy is independent of the number of scales used. Thus complex approximately analytic wavelets provide an excellent alternative to Gabor wavelets with the potential to overcome the above mentioned shortcomings of the Gabor wavelets [13]. Celik *et al.* [14] used the DT-CWT and Gabor wavelets for facial feature extraction, where they report comparable performance of the DT-CWT with more efficient computational complexity. In [15] and [16] Sun and Du applied DT-CWT for face detection and face recognition respectively.

In this paper we systematically study complex wavelets for the face recognition problem. Specifically we employ the recently developed dual-tree complex wavelet transform and a new single-tree complex wavelet transform with improved shift invariance and directional selectivity properties. First Gabor wavelet and complex wavelet based representations of face images are obtained. For all the transforms the representations encompass 4 levels and 6 directions. PCA [2, 3] and LDA [4, 5] are employed to further reduce the dimensionality of the derived feature vectors. Results of experiments carried out on FERET and ORL data-bases indicate that complex wavelets indeed constitute an excellent alternative to Gabor wavelets in face image representation and recognition.

2 GABOR Wavelets

A Gabor wavelet filter is a Gaussian kernel function modulated by a sinusoidal plane wave:

$$\Psi_g(x, y) = \frac{f^2}{\eta\gamma\pi} \exp(\beta^2 y'^2 - \alpha^2 x'^2) \exp(2\pi j f x') \quad (1)$$

$$x' = x \cos \theta + y \sin \theta, y' = y \cos \theta - x \sin \theta$$

where f is the central frequency of the sinusoidal plane wave, θ is the anti-clockwise rotation of the Gaussian and the envelope wave, α is the sharpness of the Gaussian along the major axis parallel to the wave, and β is the

sharpness of the Gaussian minor axis perpendicular to the wave. $\gamma = f/\alpha$ and $\eta = f/\beta$ are defined to keep the ratio between frequency and sharpness constant [6].

Gabor wavelets possess many properties which make them attractive for many applications. Directional selectivity is one of the most important of these properties. The Gabor wavelets can be oriented to have excellent selectivity in any desired direction. They respond strongly to image features which are aligned in the same direction and their response to other feature directions are weak. Invariance properties to shifts and rotations also play an important role in their success. In order to accurately capture local features in face images a space frequency analysis is desirable. Gabor functions provide the best tradeoff between spatial resolution and frequency resolution. The optimal frequency-space localization property allows Gabor wavelets to extract the maximum amount of information from local image regions. This optimal local representation of Gabor wavelets makes them insensitive and robust to facial expression changes in face recognition applications. The representation is also insensitive to illumination variations due to the fact that it lacks the DC component. Last but not least there is a strong biological relevance of processing images by Gabor wavelets as they have similar shapes to the respective fields of simple cells in the primary visual cortex.

Despite many advantages of Gabor wavelet based algorithms in face recognition, the high computational complexity and high memory capacity requirement are important disadvantages. With a face image of size 128×128 the dimension of the extracted Gabor features would be 655,360 when 40 wavelets are used. This feature is formed by concatenating the result of convolving the face image with all the 40 wavelets. Such vector dimensions are extremely large and, in most cases, downsampling is employed before further dimensionality reduction techniques such as PCA is applied. The computational complexity is high even when fast Fourier transform (FFT) is employed.

Because of the above mentioned shortcomings one usually looks for other transforms that can preserve most of the desired properties of Gabor wavelets and at the same time reduces the computational complexity and memory requirement. Complex wavelet transforms provide a satisfactory alternative to this problem.

3 Complex Wavelet Transform

3.1 Dual-Tree Complex Wavelet Transform

One of the most promising decomposition that removes the above drawbacks satisfactorily is the dual-tree complex wavelet transform (DT-CWT) [17, 18, 19]. Two classical wavelet trees (with real filters) are developed in parallel, with the wavelets forming (approximate) Hilbert pairs. One can then interpret the wavelets in the two trees of the DT-CWT as the real and imaginary parts of some complex wavelet $\Psi_c(t)$. The requirement for the dual-tree setting for forming Hilbert transform pairs is the well-known half-sample delay

condition. The resulting complex wavelet is then approximately analytic (i.e., approximately one sided in the frequency domain).

The transform has the ability to differentiate positive and negative frequencies and produces six sub-bands oriented in $\pm 15^\circ$, $\pm 45^\circ$, $\pm 75^\circ$. However these directions are fixed unlike the Gabor case where the wavelets can be oriented in any desired direction.

3.2 Single-Tree Complex Wavelet Transform

Complex wavelets with improved analytic property (better suppression of the negative frequencies) are possible in the single tree context. With improved analyticity property the wavelets become more selective and respond more strongly to the six fixed directions of the DT-CWT. Additionally as a consequence of the improved analyticity shift invariance property of the wavelets also improves. Thus it becomes possible to design wavelets which can imitate Gabor wavelets more closely. Complex wavelets with desired properties such as symmetry and orthogonality have been extensively studied in the literature [20, 21, 22]. These wavelets however are not analytic and thus do not possess the properties associated with analytic wavelets. We now describe the construction of approximately analytic complex wavelet transforms which possess all the properties of the DT-CWT with better directional selectivity and better shift invariance properties.

Fig. 2 shows an example of applying these transformations to a face image is shown, where the magnitude of the transformation is shown.

4 Proposed Method

In order to alleviate the computational burden and high memory requirement of the Gabor wavelet based face recognition, and at the same time retain most of its desired properties, we propose to use complex, approximately analytic wavelets instead of Gabor wavelets. We specifically consider two alternatives; the complex dual-tree wavelet transform and the complex single-tree wavelet transform described in section 3. For both approaches, the directional multiscales decomposition of the gray level face image is performed up to level 4. The DT-CWT or ST-CWT feature vector X is formed by concatenating the results of the multiscale representation. Given an image $I(x, y)$ and a wavelet $\Psi_{\mu, \nu}(x, y)$, of level μ and direction ν , vector X can be formed by:

$$X = (O_{0,0}O_{0,1} \dots O_{3,5})^t \quad (2)$$

where $O_{\mu, \nu}(x, y) = I(x, y) * \Psi_{\mu, \nu}(x, y)$ and $O_{\mu, \nu}$ $\mu = 0, \dots, 3$, $\nu = 0, 1, \dots, 5$ is formed by concatenating the rows or columns of $O_{\mu, \nu}(x, y)$. Here $*$ and t denote, the convolution and transpose operators respectively. This representation encompasses different scales, spatial location and 6 fixed directions similar to Gabor representation.

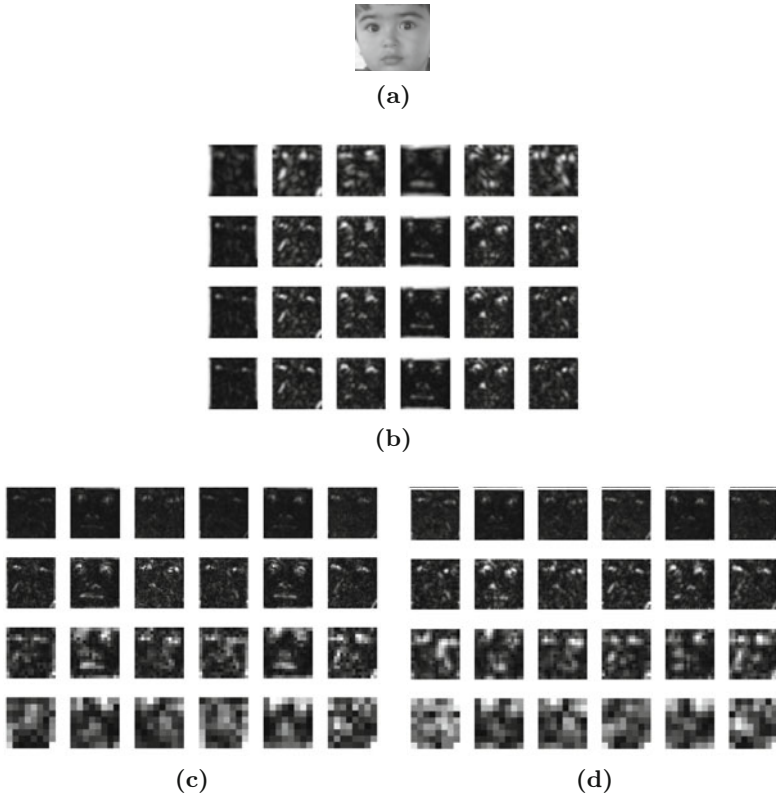


Fig. 1. (a) Sample Face image, Magnitude of transformation of a sample image using (b) Gabor wavelets, (c) DT-CWT and (d) ST-CWT with 4 scales and 6 orientations

The size of such a feature vector is 32640 pixels which is much smaller than the corresponding Gabor feature vector where the size is 393216. To reduce the dimensionality of the feature vector space we employed PCA and LDA on the Gabor, DT-CWT, and ST-CWT feature vectors. Fig. 2 shows the block diagram of the proposed method. The similarity measures used in our experiments to evaluate the efficiency of different representation and recognition methods include L_1 distance measure, δ_{L_1} , L_2 distance measure, δ_{L_2} , and cosine similarity measure, δ_{\cos} . The measures for n dimensional vectors are defined as follows:

$$\delta_{L_1}(x, y) = |x - y| \quad (3)$$

$$\delta_{L_2}(x, y) = \|x - y\|^2 \quad (4)$$

$$\delta_{\cos}(x, y) = -\frac{xy}{\|x\| \|y\|} \quad (5)$$

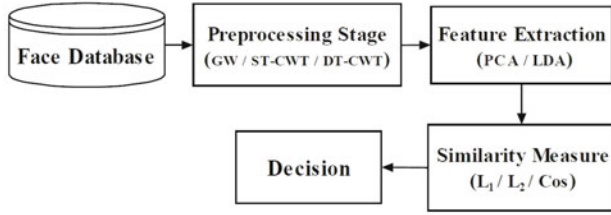


Fig. 2. The block diagram of the proposed method

We conducted experiments on two commonly used face databases: FERET and ORL data-bases. For FERET database [23], 600 frontal face images from 200 subjects are selected, where all the subjects are in an upright, frontal position. The 600 face images were acquired under varying illumination conditions and facial expressions. Each face image is cropped to the size of 128×128 and normalized to zero mean and unit variance. To test the algorithms, two images of each subject are randomly chosen for training, while the remaining one is used for testing (i.e. 400 training and 200 test images).

The ORL database [24] consists of 400 images acquired from 40 persons with variations in facial expression and facial details. All images in the database are resized to 128×128 pixels for our experiments. Out of the 10 images per subject of the ORL face database, first 5 were selected for training and the remaining 5 were used for testing (i.e. 200 training and 200 test images). Hence, no overlap exists between the training and test face images.

5 Simulation Results and Discussions

We used the derived features from Gabor, DT-CWT and ST-CWT together with PCA and LDA as a dimensionality reduction techniques to asses the performance of the complex wavelet based representations.

Table I shows the face recognition performance of PCA, Gabor+PCA, DT-CWT+PCA and STCWT+PCA for the three similarity measures using the FERET and ORL databases. For the FERET database PCA applied on raw face images recorded a recognition rate which was always less than 79%. The performance of the Gabor+PCA, DT-CWT+PCA and ST-CWT+PCA is significantly better than that of raw PCA. These results indicates that CWT based features are not as sensitive as PCA to illumination variations and facial expression changes. Moreover, ST-CWT+PCA outperformed DT-CWT+PCA and both recorded performance better than Gabor+PCA.

Similar results hold for the ORL database. PCA applied on raw face images recorded a recognition rate which was always less than 92%. The performance of the Gabor+PCA, DT-CWT+PCA and ST-CWT+PCA is again significantly better than that of raw PCA. When all features are employed

the highest recognition rates recorded for Gabor+PCA, DT-CWT+PCA and ST-CWT+PCA are respectively 93.59%, 94.75% and 94.92%.

In same manner as in Table 1, Table 2 summarizes the results for the same conducted experiments but with PCA replaced by LDA for dimensionality reduction. The same conclusions hold here with generally better performances than that of Table 1 due to the use of LDA. We conclude that complex wavelet representation based face recognition performs slightly better than of Gabor wavelet representation and the ST-CWT does slightly better than DT-CWT for both databases using PCA and LDA algorithms.

Table 1. Face recognition performance for different approaches using ORL/FERET databases with three different similarity measures using PCA for dimensionality reduction

<i>Approach</i>	δ_{L_1}	δ_{L_2}	δ_{\cos}
PCA	88.25/77.33	91.0/78.17	91.75/78.5
Gabor+PCA	93.1/92.83	92.5/91.17	93.59/91.17
DT-CWT+PCA	94.1/93.00	94.0/89.83	94.75/91.17
ST-CWT+PCA	94.59/93.33	94.33/91.83	94.92/92.0

Table 2. Face recognition performance for different approaches using ORL/FERET databases with three different similarity measures and LDA for dimensionality reduction

<i>Approach</i>	δ_{L_1}	δ_{L_2}	δ_{\cos}
LDA	90.83/80.17	94.33/81.5	92.56/80.83
Gabor+LDA	95.17/94.56	94.76/92.33	95.38/92.67
DT-CWT+LDA	96.36/95.50	96.13/91.76	97.05/93.67
ST-CWT+LDA	96.95/95.83	96.83/93.17	96.83/94.33

6 Conclusion

Complex wavelets possess most of the properties of Gabor wavelets such as good directional selectivity and invariance to shifts and in plane rotations and a representation that is local. They, however, have important advantages: they are computationally much more efficient and enjoy a much less redundant representation.

PCA and LDA are employed to further reduce the dimensionality of the complex wavelet based feature space. In all experiments carried out on standard databases, complex wavelets performed equally well or suppressed the performance of Gabor wavelets in recognition rate when equal number of orientations and scales are used. Moreover, the single tree complex wavelet transform recorded better recognition rates than that of dual tree complex

wavelet transform. These findings indicate that complex wavelets can provide a successful alternative to Gabor wavelets for face recognition. Furthermore it indicates that ST-CWT performs slightly better than DT-CWT due to the fact that it has improved directional selectivity and shift invariance properties.

References

1. Chellappa, R., Wilson, C.L., Sirohey, S.: Human and machine recognition of faces: a survey. *Proceedings of the IEEE* 83(5), 705–741 (1995)
2. Turk, M., Pentland, A.: Eigenfaces for recognition. *Journal of Cognitive Neuroscience* 3(1), 71–86 (1991)
3. Kirby, M., Sirovich, L.: Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 12(1), 103–108 (1990)
4. Zhao, W., Chellappa, R., Nandhakumarm, N.: Empirical performance analysis of linear discriminant classifiers. In: *Proceedings of Computer Vision and Pattern Recognition*, pp. 164–169 (1998)
5. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7), 771–780 (1997)
6. Shen, L., Bai, L., Fairhurst, M.: Gabor wavelets and general discriminant analysis for face identification and verification. *Pattern recognition Letters* 25(5), 553–563 (2007)
7. Lee, C.J., Wang, S.D.: Fingerprint feature extraction using Gabor filters. *Electronics Letters* 35(4), 288–290 (1999)
8. Wang, X., Ding, X., Liu, C.: Optimized Gabor filter based feature extraction for character recognition. In: *Proceeding of 16th International Conference on Pattern Recognition*, vol. 4, pp. 223–226 (2002)
9. Lades, M., Vorbruggen, J.C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R.P., Konen, W.: Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers* 42(3), 300–311 (1993)
10. Wiskott, L., Fellous, J.M., Kruger, N., von der Malsburg, C.: Face recognition by elastic bunch graph matching. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 19(7), 775–779 (1997)
11. Liu, D.H., Lam, K.M., Shen, L.S.: Optimal sampling of Gabor features for face recognition. *Pattern Recognition Letters* 25(2), 267–276 (2004)
12. Liu, C., Wechsler, H.: Independent component analysis of Gabor features for face recognition. *IEEE Transaction on Neural Networks* 14(4), 919–928 (2003)
13. Eleyan, A., Ozkaramanli, H., Demirel, H.: Complex wavelet transform-based face recognition. *EURASIP Journal on Advances in Signal Processing*, 1–13 (2008)
14. Celik, T., Ozkaramanli, H., Demirel, H.: Facial feature extraction using complex dual-tree wavelet transform. *Computer Vision and Image Understanding* 111(2), 229–246 (2008)
15. Sun, Y., Du, M.: Face detection using DT-CWT on SHPCA space. In: *Proceeding of 6th International Conference on Intelligent Systems Design and Applications*, vol. 2, pp. 433–437 (2006)

16. Sun, Y., Du, M.: DT-CWT feature based classification using orthogonal neighborhood preserving projections for face recognition. In: Proceeding of International Conference on Computational Intelligence and Security, pp. 719–724 (2006)
17. Kingsbury, N.G.: The dual-tree complex wavelet transform: a new efficient tool for image restoration and enhancement. In: Proceedings of 9th European Signal Processing Conference, pp. 319–322 (1998)
18. Kingsbury, N.G.: Shift invariant properties of the dual-tree complex wavelet transform. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 3, pp. 1221–1224 (1999)
19. Selesnick, I.W., Baraniuk, R.G., Kingsbury, N.G.: The dual-tree complex wavelet transform-A coherent framework for multiscale signal and image processing. *IEEE Signal Processing Magazine* 22(6), 123–151 (2005)
20. Zhang, X.P., Desai, M.D., Peng, Y.N.: Orthogonal complex filter banks and wavelets: Some properties and design. *IEEE Transactions on Signal Processing* 47(4), 1039–1048 (1999)
21. Gao, X.Q., Nguyen, T.Q., Strang, G.: A study of two-channel complex-valued filterbanks and wavelets with orthogonality and symmetry properties. *IEEE Transactions on Signal Processing* 50(4), 824–833 (2002)
22. Lina, J.M., Mayrand, M.: Complex Daubechies wavelets. *Applied and Computational Harmonic Analysis* 2(3), 219–229 (1995)
23. Philipps, P.J., Moon, H., Rivzi, S., Ross, P.: The Feret evaluation methodology pro face-recognition algorithms. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 22(10), 1090–1100 (2000)
24. Samaria, F., Harter, A.: Parameterization of A Stochastic Model for Human Face Identification. In: Proceedings of 2nd IEEE Workshop on Applications of Computer Vision, pp. 138–142 (1994)

Biomedical Image Processing

The Method of Immunohistochemical Images Standardization

Anna Korzyńska¹, Urszula Neuman¹, Carlos Lopez²,
Marylene Lejeun², and Ramon Bosch³

¹ Nalecz Institute of Biocybernetics and Biomedical Engineering,
Polish Academy of Sciences,
ul. Księcia Trojdena 4, 02-109 Warszawa, Poland
e-mail: anna.korzynska@ibib.waw.pl

² Molecular Biology and Research Section,
Hospital de Tortosa Verge de la Cinta, C/Esplanetes no 14,
Tortosa 43500, Tarragona, Spain

³ Department of Pathology, Hospital de Tortosa Verge de la Cinta,
C/Esplanetes no 14, Tortosa 43500, Tarragona, Spain

Summary. The standardization method of immunohistochemically staining tissue section images prior to the image processing and analysis is described in this paper. The effectiveness of the proposed standardization method is examined on thin tissue slices of breast cancer stained with DAB & H. The image analysis results after the initial image standardization are more closer to the results of traditional methods of cells nuclei quantification than for original images.

1 Introduction

Nowadays developments of computers and microscopes technology and its lower costs cause the computer-assisted acquisition of histochemically and immunohistochemically staining tissue section images are possible in any hospital. The analysis and quantification of such images usually is performed by experienced pathologist according to sterology methodology. Visual human analysis and scoring suffers from intra- and interobserver variability and from uncertainty of cut off levels. Automatic analysis of this type of images is not easy because of its huge variability in image overall and local contrasts and variability of color ranges observed in images. This variability is caused by biological variability of pathological tissue itself [1], various steps of samples preparing procedure [2, 3] acquisition processes [6]. So, at the beginning of the 1990s [5], telepathology was established as an idea of the examination of microscopic images from a distance, using image processing and analysis methods. The telepathology supports physicians and family doctors out of university centers in computer-assisted diagnosis, grading and prognosis based on images captured and send to academic centers. It increase unification of a criteria of quantitative analysis.

First systems for microscopic image analysis in histopathology, e.g. iPATH or UICC-TPCC [6], were established at the beginning of this century. Then, commercially available systems for digitalization of complete glass slides as well as internet-based automated image measurement systems, such as EAMUSZ [4], followed them. There are some complex and sophisticated algorithms [7, 8, 9, 10, 11, 12, 13], which have been tested for various types of histopathological measurements. However, they all lack of universality and, in a case of changes of image characteristics, they produce moderate results [14, 15, 6].

Systems of telepathology consist of image transfer, image segmentation, objects measurement and/or counting, statistical analysis parts [5]. Results of specific type of images analysis and quantification, obtained from a system, are sent back to the recipients, supporting physicians in their decisions.

Information automatically extracted from images is reproducible and reliable and of high-quality in the sense of "biological" content if samples were prepared according to medical protocols and if images shows representative areas of samples and if adjustment of microscope, its illumination conditions, quality and chosen settings of the camera are the best of possible. The authors of this paper do not have any influence of most of these factors because images used in investigation were acquired by various pathologists using various devices. But images acquired by one pathology using one set of capturing devices also varies in color tone, brightness, white balance and contrast. So the aim of this investigations is to unify set of images on the level of it performance to make them more suitable for automatic analysis. In this step of analysis the proposed method of image standardization is limited to deal with set of images acquired using one set of capturing devices by the same person.

2 Images Characteristics

Images used in this investigation were captured by microscope connected with digital camera (described farther) by one pathologist. They show various breast cancer patients samples which were indirectly stained against KI-67 protein (an nuclear marker of cell prolypheration) used in evaluation of patient prognosis. The end product of staining procedure 3,3'-diaminobenzidine (DAB) allows to distinguish cells nuclei with positive immunoreactiveness (brown) , from others with negative immunoreactiveness, contrastained with hematoxylin (H) (blue) [16, 17]. So images show blue and brown color rounded objects - cells' nuclei - distributed in various homogeneity across the image plane. In some parts of an image the single brown objects are surrounded by blue ones, while in other parts the clusters of brown objects, overlapping or touching one to another, are placed. It is observed that there is a variation in blue and brown colors in one image and from one image to another, see Fig. 1. In single image also non-homogeneity of light distribution is observed: meddle part of image is brighter then its peripheral part, what is visible in

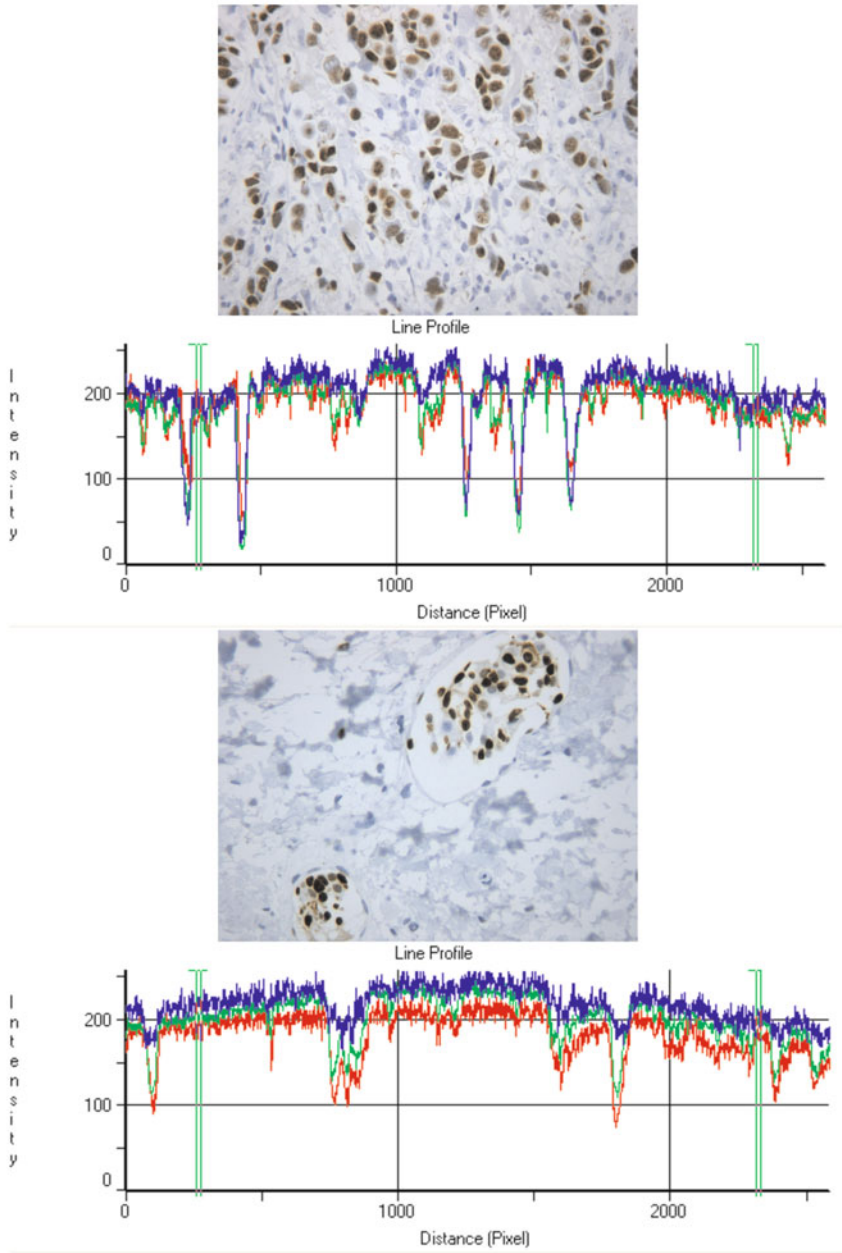


Fig. 1. Examples of immunohistochemically stained breast cancer tissue section images with line profiles along diagonal line from upper left to bottom right corners

Fig. 1. It was found that even images collected by one pathologist using particular microscope and camera devices are different from the others because of variation in color and contrast. It can be caused by changes in external light conditions, changes in chosen parameters of image acquisition according to sample density. Variability of images contrast and range of its blue and brown color should be decreased to achieve homogeneity on such level that automatic analysis results will agree with human perception.

3 Methods

The standardization method of digital color images of immunohistochemically stained tissue section of breast cancer acquired using certain acquisition devices is proposed. The proposed method of image color standardization consists of following steps: (1) estimation of characteristic of microscopy and camera light distribution across the image plane constructing so called correction map, (2) single image homogeneity in lightness distribution correction with this map, (3) single image contrast rearrangement by manipulation of range of gray levels histogram of channel Value in HSV color model. Map of light distribution is constructed based on about hundred images of low object density as a median of each point across image plane, while homogeneity correction is done by subtracting correction image (transferred version of correction map) from original image. Contrast enhancement is done by linear histogram transform with small saturation of its black part.

Results of each step of the procedure of standardization is shown in Fig. 2. Only V - channel of HSV color model is presented. The human visual system finds color images after proposed transformation more clear, with increasing local and overall contrast and with objects darker than in original image.

Since there is not any definition of immunohistochemically stained images quality and it is hard to define a measure of a distance between image quality, the effectiveness of proposed method is examined in this paper based on the set of carefully acquired images using two reference systems. Firstly it is the comparison of the number of brown objects selected with and without of the step of color standardization by chosen automated segmentation method and secondly the comparison with objects detection done by pathologist.

To evaluate proposed image standardization method the method of segmentation and counting objects in DAB & H stained images of tissue section was used. This segmentation method was developed as a modification of segmentation and counting procedure proposed early in [16] for lymphoma tissue section images stained immunohistochemically against FOXP3 with DAB & H. As it was published the results of counting T-cells with positive immunoreactivity are in good agreement with the results of counting manually by marking objects [16]. But for breast cancer tissue section images the method was improved by changing adaptive threshold into several levels of global threshold.

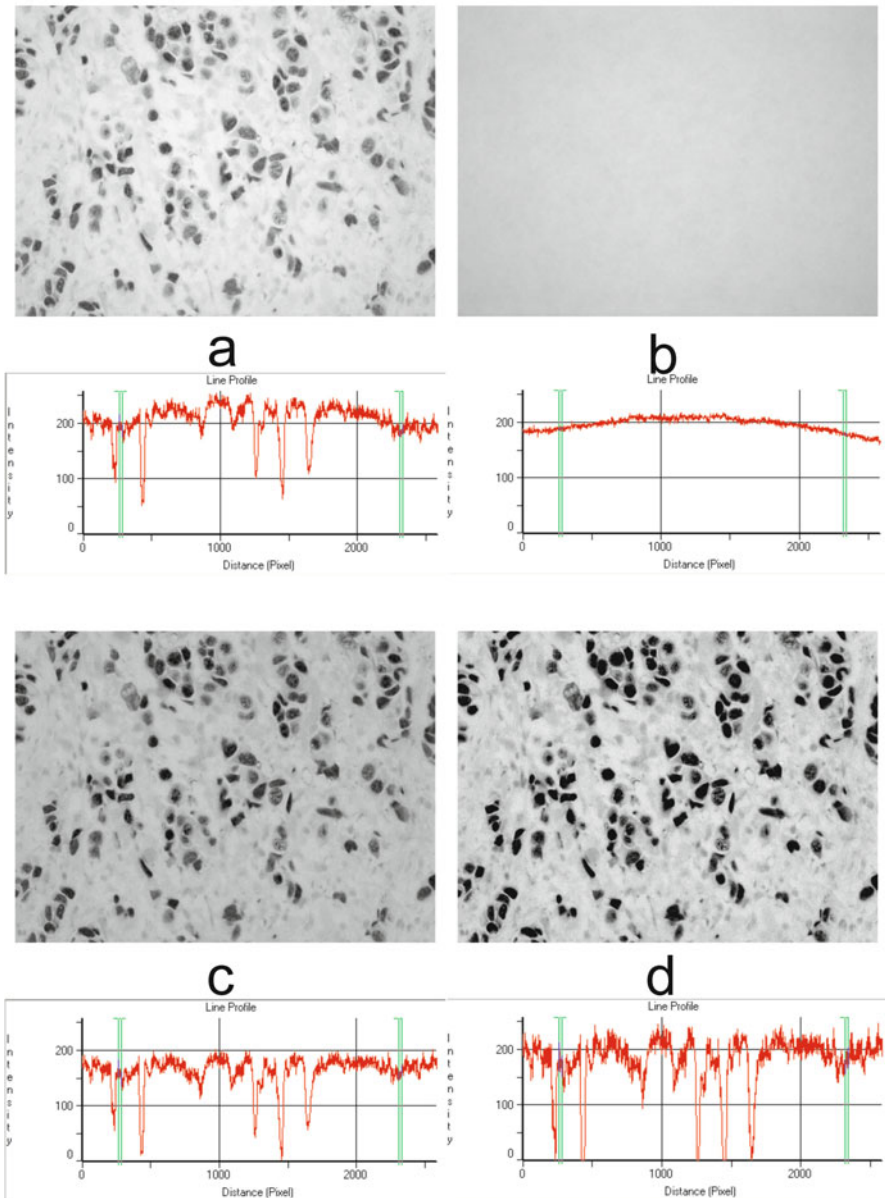


Fig. 2. Process of image standardization: (a) V - channel of HSV model of colors from image presented in upper part of Fig. 1, (b) Map of light distribution in acquiring devices described in section Experimental Data (1), (c) V - channel after illumination correction (2), (d) V - channel after contrast increment (3)

4 Experimental Data

The set of 60 digital images were acquired in Hospital de Tortosa Verge de la Cinta based on the collection of paraffin-embedded, breast cancer tissue sections, coming from surgical interventions. 3 μm thick sections were dried, deparaffinized in xylene, rehydrated in graded ethanol, and washed in water and PBS. Antigen retrieval was achieved by heat treatment in a pressure cooker. Samples were incubated with the appropriate dilution of the primary antibodies: Ki-67 (clone MIB-1, Dako, Carpinteria, CA, USA). Automatic immunostaining was performed with a Horizon TechMate (Dako). The secondary antibody was conjugated to the reagents in the Dako EnVision⁺ system (Dako EnVision⁺, Dako Corporation), which includes peroxidase block, labeled polymer, and buffered substrate/DAB⁺ chromogen. Targeting antigen DAB forms a brown end product. Finally, tissues were counterstained with hematoxylin, dehydrated, and mounted. This process has been standardized to ensure stain homogeneity. Prepared samples were used to acquire digital images of tissue section by bright field techniques (Leica DM LB2 upright light microscope, Leica Microsystems Wetzlar GmbH, Wetzlar, Germany) with 40x plane-apochromatic objective of numerical aperture 0,63. According to experienced pathologist choice 60 areas of various complexity of the several samples were captured using Leica DFC320 digital camera (Leica Microsystems Digital Imaging, Cambridge, UK) controlled by Leica DFC Twain software 6.3.0 (Leica Microsystems Digital Imaging). During this process stable illumination conduction were carefully organized. Each collected image has covered about 11 mm² of samples area with pixel resolution 2088 \times 1550 and have been registered as RGB 24-bits color image then stored in uncompressed file, examples of this images are shown in Fig. [11](#).

5 Results and Discussion

It has been observed that most of acquired images seems to be similar in the sense of human visual system examination. But 3 of them have been picked by pathologists as images of outstanding characteristics. For all images the number of cells' nuclei with positive immunoreactivity was counted by an experienced pathologist, by automatic method with and without preprocessing of the proposed image standardization procedure.

For each image the proposed standardization procedure has changed its performance in the way that human visual system perceps them as more clear and better contrasted. Automatic segmentation and counting procedure improves results of object counting in relation to 'gold standard' results of manual counting when it is applied to images after standardization. The mean and the standard deviation of the relative error in immunopositive cell counting decreases from $0,28 \pm 2,41$ to $0,59 \pm 0,95$ for 60 images (4 complicated images have been excluded from analysis because the automatic segmentation

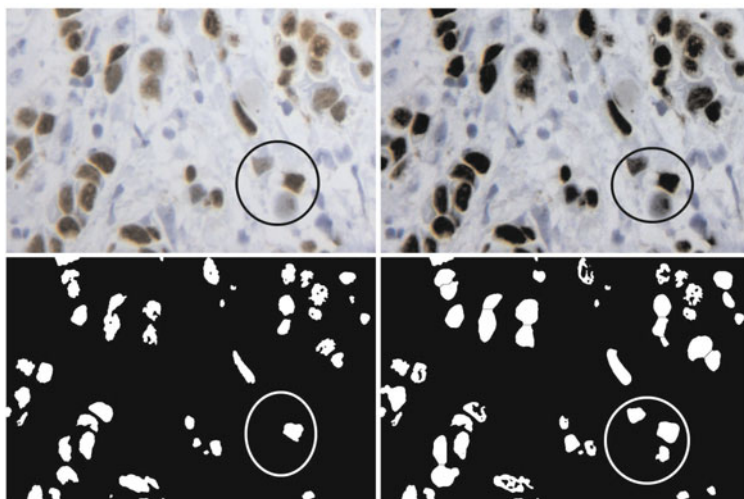


Fig. 3. Results of the segmentation and counting procedure for left bottom part of original image shown in upper line in Fig. 1 (left) and its standardized version (right)

procedure fails despite the fact that image standardization helps segmentation procedure to produce better results than without).

6 Conclusion

The results show that the proposed method of color and contrast standardization is useful in supporting the automatic image analysis. We found images collected by one pathologist using the same acquiring devices not homogeneous in light color and contrast distribution because of variation in sample density and variation in light conditions during acquisition. The proposed method of image standardization reduces this variability so image analysis results after the initial image standardization are more similar to the results of traditional methods of quantifying the number of cells nuclei with positive immunoreactivity than without preprocessing. In further work tests of proposed standardization method with images captured using various devices but showing close view plane of some samples will be performed. These new tests allow to develop and adjust proposed methodology to the interobserver and various devices variability.

Acknowledgement

This study was supported by different Short Term Scientific Mission (STSM) of the COST Action IC0604 Euro-Telepath that permitted collaborative stud-

ies between the Nalecz Institute of Biocybernetics and Biomedical Engineering (Poland) and the Hospital de Tortosa Verge de la Cinta (Spain): STSM-5044 (A. Korzyńska), STSM-5045 (U. Neuman), STSM-5043 (M. Lejeune) and STSM-5046 (R. Bosch).

References

1. Swerdlow, S., Campo, E., Harris, N., Jaffe, E., Pileri, S., et al.: WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. In: IARC (2007)
2. Seidal, T., Balaton, A.J., Battifora, H.: Interpretation and quantification of immunostains. *The American Journal of Surgical Pathology* 25(9), 1204–1207 (2001)
3. Yaziji, H., Barry, T.: Diagnostic immunohistochemistry: what can go wrong? *Adv. Anat. Pathol.* 13(5), 238–246 (2006)
4. Kayser, G., Radziszowski, D., Bzdyl, P., Sommer, R., Kayser, K.: Theory and implementation of an electronic, automated measurement system for images obtained from immunohistochemically stained slides. *Anal. Quant. Cytol. Histol.* 28(1), 27–38 (2006)
5. <http://thedoctorsdoctor.com>
6. Kayser, K., Radziszowski, D., Bzdyl, P., Sommer, R., Kayser, G.: Towards an automated virtual slide screening: theoretical considerations and practical experiences of automated tissue-based virtual diagnosis to be implemented in the internet. *Diagnostic Pathology* 1(1), 10 (2006)
7. Hyun-Ju, C., Ik-Hwan, C., et al.: Color image analysis for quantifying renal tumor angiogenesis. *Analyt. Quant. Cytol. Histol.* 27, 43–51 (2005)
8. Bartels, P., Montironi, R., Duval da Silva, V., Hamilton, P., Thompson, D., et al.: Tissue architecture analysis in prostate cancer and its precursors: An innovative approach to computerized histometry. *Rur. Urol.* 35, 484–491 (1999)
9. Schulerud, H., Kristensen, G., Liestol, K., Vlatkovic, L., Reith, A., Albbregtsen, F., Danielsen, H.: A review of caveats in statistical nuclear image analysis. *Analit. Cell. Pathol.* 16, 63–82 (1998)
10. Kan, J., Qing-Min, L., Sheng-Yang, D.: A novel white blood cell segmentation scheme using scale-space filtering and watershed clustering. In: 2003 International Conference Machine Learning and Cybernetics, vol. 5, pp. 2820–2825 (2003)
11. Markiewicz, T., Wiśniewski, P., Osowski, S., Patera, J., Kozłowski, W., Koktysz, R.: Comparative analysis of methods for accurate recognition of cells through nuclei staining of KI-67 in neuroblastoma and estrogen/progesterone status staining in breast cancer. *Analyt. Quant. Cytol. Histol.* 31(1), 49–62 (2009)
12. Kóprowski, R., Wróbel, Z.: Automatic segmentation of biological cell structures based on conditional opening or closing. *MG&V* 14, 285–307 (2005)
13. Markiewicz, T., Osowski, S., Pater, J., Kozłowski, W.: Image processing for accurate cell recognition and count on histologic slides. *Analyt. Quant. Cytol. Histol.* 28(5), 281–291 (2006)
14. Pavlopoulos, P., Zimeras, S., Kavantzias, N., Korkolopoulou, P., et al.: Segmentation of transitional cell carcinoma nuclei by unsupervised thresholding in different color spaces. *Anal. Quant. Cytol. Histol.* 29(4), 271–278 (2007)

15. Wang, M., Zhou, X., Li, F., Huckins, J., King, R., Wong, S.: Novel cell segmentation and online svm for cell cycle phase identification in automated microscopy. *Bioinformatics* 24(1), 94–101 (2008)
16. Neuman, U., Korzyńska, A., Lopez, C., Lejeun, M.: Segmentation of stained lymphoma tissue section images. In: Pietka, E., Kawa, J. (eds.) *Information Tech. and Biomedicine*, ASC, Springer, Heidelberg (2010) (accepted)
17. Sung-Hyuk, C.: A fast hue-based colour image indexing algorithm. *MG&V* 11(2/3), 285–295 (2002)
18. Fu, K., Mui, J.: A survey on image segmentation. *Pattern Recognition* 13(1), 3–16 (1981)
19. Pham, D., Xu, C., Prince, J.: A survey of current methods in medical image segmentation. *Annual Review of Biomedical Engineering* 2, 315–338 (2000)
20. Iwanowski, M., Soille, P.: Morphological Refinement of an Image Segmentation. LNCS, pp. 538–545. Springer, Heidelberg (2005)

The Usefulness of Textural Features in Prostate Cancer Diagnosis

Jacek Śmietański

Institute of Computer Science, Jagiellonian University,
ul. Łojasiewicza 6, 30-348 Kraków,
e-mail: jacek.smietanski@ii.uj.edu.pl

Summary. To enable an effective treatment, the prostate cancer (PCa) must be detected early enough. Unfortunately, the diagnostic methods are insufficient. The hope for improve the PCa diagnosis lies in the perfusion computed tomography (p-CT) method. However, the p-CT prostate images are not easy to interpret.

The presented work describes the technique of computational analysis of such images using the textural features of the Haralick's co-occurrence matrices. The research based on the material from over 50 patients concentrated on selection of proper preprocessing procedures, optimal feature space and the best decision function. A serious problem was also to choose regions of interest - especially important areas in the gland.

It seems that the improvement of detectability of PCa with the p-CT technology is possible by creating a dedicated computational system to CT scanners, that could point out the cancerous lesions automatically, faster, and more reliable than in traditional methods.

1 Prostate Cancer Diagnostics

The prostate cancer (PCa) is one of the most common malignant cancers. For example in the United States it accounts for 25% of new cancers and 10% of cancer deaths among American men [1]. The American Cancer Society estimates that by the end of 2009, there were 192 280 new cases and 27 360 deaths attributed to PCa [1]. Fortunately, the PCa can be successfully treated. However, it is possible only if detected early enough, before the metastasis occurs. The earlier stage and grade, the better chance to effective treatment. Therefore the early detection of PCa is a key problem.

There are many techniques which enable us to expect the PCa existence. The most popular are the digital rectal examination (DRE) and PSA (prostate specific antigen) measure. The first one may be helpful only if cancer is localized on the surface of the gland, near rectum, the second one is not specific enough as the PSA level naturally increases with the age and there is no strictly defined level which can separate cases with the PCa and the other. [10, 12, 13].

The only method which can certainly confirm the PCa existence, is biopsy. However, biopsy may be successful only if the biopsy needle hits precise into the cancerous lesion. If the cancer is small, also the probability of its detection is in low level. The biopsy is often conducted under the transrectal ultrasonography (TRUS) control. However, also in TRUS the cancerous lesions are often not visible. [6, 11]

2 Perfusion Computed Tomography

In view of this the need of new, more effective method is obvious. One of techniques which may improve the effectiveness of PCa diagnosis is perfusion computed tomography (p-CT). In convectional CT, the early PCa is not visible. However, the perfusion option enables us to see not only the anatomical structure but also the functioning of the diagnosed organs. In this method the perfusion values like blood flow (BF), blood volume (BV), mean transit time (MTT) and permeability surface (PS) are measured. The meaning and method of calculation those parameters are described in [9, 15]. In our experiments they were calculated automatically by the *GE Advanced Workstation* [5], which is the part of the multislice CT scanner used for diagnosing patients.

It is supposed that - due to the angiogenesis caused by growing cancer [4, 9], when the PCa occurs, the BF, BV and PS rise, and MTT decreases in the cancerous lesion. However, the prostate is a very little blooded gland and the differences of perfusion parameters may be too small for precise measurement and point out the proper region. This implies difficulties in manual interpretation of p-CT prostate images. There are no clearly visible objects or clearly distinguishable areas on them. Therefore the author decided to use the computational techniques for such p-CT images interpretation.

The computational analysis has three main advantages. First - we are able to extract and analyze many features which are invisible for human observer. Second - computational analysis is deterministic and independent of individual properties of eyesight. And finally - the analysis may be faster, cheaper and more reliable. To interpret the given images the statistical textural features derived from co-occurrence matrices were used.

3 The p-CT Images

The images from 56 patients with suspected PCa (because of high PSA level) and from one healthy patient (without PCa) were diagnosed. For each patient the p-CT examination was conducted. The scanning was started about 10s after administration of 50ml non-ionic contrast medium (370mgI/ml) at the rate 5ml/s, and lasted 50s. The total width of the diagnosed area was 20mm. The results were calculated on three levels, conventionally base, middle and apex of the prostate. For each level four images - parametric maps (BF, BV, MTT and PS) - were constructed (Fig. 1).

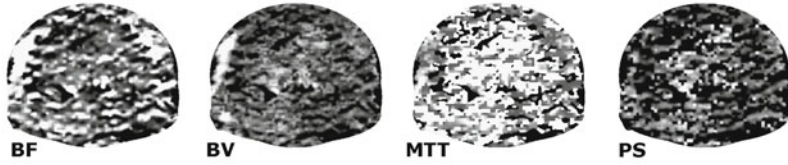


Fig. 1. p-CT images of the prostate: BF - blood flow; BV - blood volume; MTT - mean transit time; PS - permeability surface

4 Textural Features

To ensure the possibility of comparing images from different patients, they were scaled in size and had equalized histogram or the grayscale values. The analysis was based on the Haralick’s co-occurrence matrices (GLCM) [7] (Fig.2) and 21 coefficients derived from them (Tab.1).

Let $I : Z^2 \supset D \rightarrow G = \{1, \dots, N_g\}$ (where Z denotes set of integers) be a two-dimensional discrete image with N_g gray levels. For the given image I we define the GLCM:

$$P(i, j|d, \theta) = \frac{\#\{k, l \in D : I(k) = i, I(l) = j, \|k - l\| = d, \angle(k - l) = \theta\}}{\#\{m, n \in D : \|m - n\| = d, \angle(m - n) = \theta\}} \tag{1}$$

where: $i, j \in G$ – gray levels of points k and l , respectively; $\angle(k - l)$ - the angle between vector \vec{kl} and axe \vec{OX} ; d – distance between k and l ; θ - direction of co-occurrence, $\#X$ – power (number of elements) of set X .

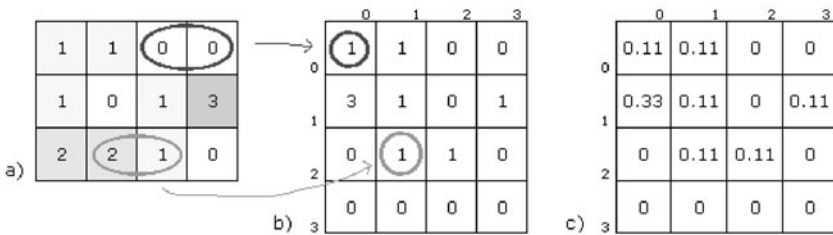


Fig. 2. Example of GLCM: a) source image with 4 gray levels; b) illustration of counting co-occurrences for $d = 1, \theta = 0^\circ$; c) GLCM, $d = 1, \theta = 0^\circ$ (counted co-occurrences are divided by number of all considered pairs of points (here 9); in this example the values were rounded to two places after comma).

Table 1. Coefficients of GLCM

no.	name	abbr.	value
f_1	energy	ENE	$f_1 = \sum_{i,j} P(i, j)^2$
f_2	entropy	ENT	$f_2 = - \sum_{i,j} P(i, j) \log P(i, j)$
f_3	homogeneity	IDM	$f_3 = \sum_{i,j} \frac{1}{1+(i-j)^2} P(i, j)$
f_4	inertia	CON	$f_4 = \sum_{i,j} (i-j)^2 P(i, j)$
f_5	correlation	COR	$f_5 = - \sum_{i,j} \frac{(i-\mu_x)(j-\mu_y)}{\sigma_x \sigma_y} P(i, j)$
f_6	variance	VAR	$f_6 = \sum_{i,j} (i+j-\mu_x-\mu_y)^2 P(i, j)$
f_7	shade	SHA	$f_7 = \sum_{i,j} (i+j-\mu_x-\mu_y)^3 P(i, j)$
f_8	prominence	PRO	$f_8 = \sum_{i,j} (i+j-\mu_x-\mu_y)^4 P(i, j)$
f_9	sum average	SA	$f_9 = \sum_{i=2}^{2N_g} iP_{x+y}(i)$
f_{10}	sum entropy	SE	$f_{10} = - \sum_{i=2}^{2N_g} P_{x+y}(i) \log P_{x+y}(i)$
f_{11}	sum variance	SV	$f_{11} = - \sum_{i=2}^{2N_g} (i-f_9)^2 P_{x+y}(i)$
f_{12}	difference average	DA	$f_{12} = \sum_{i=0}^{N_g-1} iP_{x-y}(i)$
f_{13}	difference entropy	DE	$f_{13} = - \sum_{i=0}^{N_g-1} P_{x-y}(i) \log P_{x-y}(i)$
f_{14}	difference variance	DV	$f_{14} = - \sum_{i=0}^{N_g-1} (i-f_{12})^2 P_{x-y}(i)$
f_{15}	information measure	IMC1	$f_{15} = \frac{f_2 - HXY_1}{\max(HX, HY)}$
f_{16}	coefficient of variation	COV	$f_{16} = \frac{\sigma(P(i,j))}{\mu(P(i,j))}$
f_{17}	peak transition probability	MAX	$f_{17} = \max(P(i,j))$
f_{18}	diagonal variance	DIAV	$f_{18} = \sigma^2(P(i,j))$
f_{19}	diagonal moment	DIAM	$f_{19} = \sum_{i,j} (\frac{1}{2} i-j P(i,j))^{\frac{1}{2}}$
f_{20}	second diagonal moment	DSM	$f_{20} = \sum_{i,j} \frac{1}{2} i-j P(i,j)$
f_{21}	triangular symmetry	TRS	$f_{21} = P(i,j) - P(j,i) $

Notation used in the table:

$$\mu_x = \sum_i i \sum_j P(i, j), \mu_y = \sum_j j \sum_i P(i, j),$$

$$\sigma_x = \sum_i (i - \mu_x)^2 \sum_j P(i, j), \sigma_y = \sum_j (j - \mu_y)^2 \sum_i P(i, j),$$

$$P_x(i) = \sum_j P(i, j), P_y(j) = \sum_i P(i, j),$$

$$P_{x+y}(k) = \sum_{i,j: i+j=k} P(i, j), P_{x-y}(k) = \sum_{i,j: |i-j|=k} P(i, j),$$

$$HX - \text{entropy } P_x(i), HY - \text{entropy } P_y(j), HXY_1 = - \sum_{i,j} P(i, j) \log(P_x(i)P_y(j)).$$

5 Methodology and Results

The aim was to select the set of features which used together has the best discriminatory power, and therefore are able to distinguish and point out the cancerous regions.

There was conducted the number of experiments, for each of the key 240 tests the set of 378 features was selected - the GLCM were calculated for each distance d in range 1 to 9, and for angle θ with values 0° and 90° , for each GLCM the 21 above mentioned coefficients were calculated. The resulted values for each feature were analyzed in order to eliminate outliers and normalized. The distribution of each feature was equalized using the ladder of powers method [14] (see equation 2) with $\gamma \in (0, 2]$.

$$error(\gamma) = \sum_{c=1,2} \left(\int_x [cdf\{x_c^\gamma\} - \Phi\{\overline{x_c^\gamma}, var(x_c^\gamma)\}]^2 \right) \quad (2)$$

where: $c = \{1, 2\}$ - classification; $cdf(x_c^\gamma)$, $\overline{x_c^\gamma}$, $var(x_c^\gamma)$ - distribution function, mean and variance of empirical distribution for class c , respectively; $\Phi(\mu, \sigma^2)$ - normal distribution function with mean μ and variance σ^2 .

We were looking for γ_{opt} , which minimize $error(\gamma)$ function:

$$\gamma_{opt} = \min_{\gamma} error(\gamma) \quad (3)$$

Features, where $error(\gamma_{opt}) \geq 1$, were excluded from further analysis. For each of the remaining features there was evaluated Bhattacharyya measure [3] for normal distribution:

$$J = \frac{1}{4}(\mu_1 - \mu_2)^T [\Sigma_1 + \Sigma_2]^{-1} (\mu_1 - \mu_2) + \frac{1}{2} \log \left[\frac{|\frac{1}{2}(\Sigma_1 + \Sigma_2)|}{\sqrt{|\Sigma_1||\Sigma_2|}} \right] \quad (4)$$

where: μ_1, μ_2 - means, Σ_1, Σ_2 - covariance matrices for classes 1 and 2, respectively, $|\Sigma|$ - determinant Σ .

Classification was made using the quadratic decision function:

$$d_c(x) = \log p_c - \frac{1}{2} \log |\Sigma_c| - \frac{1}{2} [(x - \mu_c) \Sigma_c^{-1} (x - \mu_c)^T] \quad (5)$$

where: $c \in 1, 2$ - index of analyzed class, p_c - probability a priori of membership of analyzed object to class c , Σ_c - covariance matrix for class c , μ_c - mean vector for class c , x - features vector for classified object.

The set of features proposed for final recognition cannot be too large because the problem of the *curse of dimensionality* [2] may occur. Therefore maximum 6-dimensional features spaces were accepted for further consideration. Another problem was the exponential complexity of the optimal algorithm for selection the best subset of 6 from the 378 features being under

consideration, which makes it useless. Finally, the suboptimal method *plus-2-take-away-1* [8] was used. In this method we start with a pair of most discriminating features. Then in each step to the final subset such two of the remaining features are added which it has the highest discriminatory power with (in this case, the highest value of the Bhattacharyya measure) and after that, the one, the least useful, feature is removed. The added features must not be correlated with the other already being in the final set.

The total number of 240 experiments were conducted - in each of them the different regions of interest selection and interpretation, different image pre-processing methods, and different perfusion parameters were used. For each experiment the best 6-dimensional feature set was selected and its usability was evaluated during the recognition process. Due to the problems with independent (based on the postoperative preparations) localization of PCa for all patients, the algorithm was tested only on 59 images selected from 159 accepted to research.

The verification based on such a control set allows to estimate the effectiveness of the proposed algorithm at 86% of correct recognitions (sensitivity 92%, specificity 67%).

6 Conclusion

It seems that the p-CT method has a big potential to recognize PCa and point out the cancerous regions. Therefore such a technique could be useful in the early PCa diagnosis, especially in cases too difficult for other techniques.

Some original algorithms were proposed and evaluated on the large set of p-CT images. The results are very promising. In the further work more clinical trials are expected. Those algorithm should be tested on new images, including more images from healthy patients. All of those images should be also interpreted by experienced pathologists to compare the computational evaluation with experts opinion.

If the results of those tests are positive, it will be possible to construct a system providing fast, automatic and deterministic identification of the cancerous lesions.

References

1. American Cancer Society: Cancer Facts and Figures, American Cancer Society, Atlanta (2009)
2. Bellman, R.E.: Adaptive Control Processes: A Guided Tour. Princeton University Press, Princeton (1961)
3. Bhattacharyya, A.: On a measure of divergence between two statistical populations defined by their probability distributions. Bulletin of the Calcutta Mathematical Society 35, 99–110 (1943)
4. Charlesworth, P., Harris, A.: Mechanisms of disease: angiogenesis in urologic malignancies. Nature Clinical Practice Urology 3(3), 157–169 (2006)

5. GE Healthcare, Advantage Workstation (2009), <http://www.gehealthcare.com/euen/advantage-workstation/index.html>
6. Halpern, E.J., Ramey, J.R., Strup, S.E., et al.: Detection of prostate carcinoma with contrast-enhanced sonography using intermittent harmonic imaging. *Cancer* 104(11), 2373–2383 (2005)
7. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics* 3, 610–621 (1973)
8. Kittler, J.: Feature set search algorithms. In: Chen, C.H. (ed.) *Pattern Recognition and Signal Processing*, Sijthoff and Noordhoff (1978)
9. Miles, K.A.: Functional computed tomography in oncology. *European Journal of Cancer* 38, 2079–2084 (2002)
10. Moul, J.W., Armstrong, A.J., Hollenbeck, B.K., et al.: Prostate Cancer. In: *Cancer Management: A Multidisciplinary Approach*, 12th edn. (2009), <http://www.cancernetwork.com/cancer-management-12/chapter14/article/10165/1536396>
11. Norberg, M., Egevad, L., Holmberg, P., et al.: The sextant protocol for ultrasound-guided core biopsies of the prostate underestimates the presence of cancer. *Urology* 50, 562–566 (1997)
12. Pelzer, A.E., Tewari, A., Bektic, J., et al.: Detection rates and biologic significance of prostate cancer with PSA less than 4.0 ng/mL: observation and clinical implications from Tyrol Screening Project. *Urology* 66, 1029–1033 (2005)
13. Thompson, I.M., Pauler, D.K., Goodman, P.J., et al.: Prevalence of prostate cancer among men with a prostate-specific antigen level \leq 4.0 ng per milliliter. *New England Journal of Medicine* 350(22), 2239–2246 (2004)
14. Tukey, J.W.: *Exploratory Data Analysis*. Addison-Wesley, Reading (1977)
15. Wintermark, M., Maeder, P., Thiran, J.P., et al.: Quantitative assessment of regional cerebral blood flows by perfusion CT studies at low injection rates: a critical review of the underlying theoretical models. *Eur. Radiol.* 11, 1220–1230 (2001)

Noise Influence Reduction in Estimation of CBF, CBV and MTT, MRI Perfusion Parameters

Rafał Henryk Kartaszyński and Paweł Mikołajczak

Maria Curie Skłodowska University,
Maria Curie-Skłodowska 1 square, 20 - 031 Lublin, Poland
e-mail: hatamoto@op.pl, mikfiz@goblin.umcs.lublin.pl

Summary. In this article we present our research into the subject of reducing the influence of noise on evaluation of perfusion parameters, such as CBF, CBV or MTT. Noise can be present on some pixels of study slices, therefore it can lead to artifacts in calculated concentration time curves and blur the final results.

To minimize influence from these factors we propose method that is different from commonly used. Generally noise reduction is done by filtering (smoothing, blurring), which is not always producing good results, as many information from image is lost. Therefore more effective is using the interpolation methods.

We have studied different interpolation techniques and compared them numerically. Tests have proven that using our method leads to better, more accurate estimation of perfusion parameters. It also seems that large window *Sinc* interpolation gives the best results.

1 Introduction

Perfusion weighted MR imaging is a technique that can provide information about functional status of cerebral tissue. It has been studied in various diseases of brain since 1989 [1, 2]. Using the signal change that brain tissue experiences over time following administration of extracellular gadolinium based contrast agents, important hemodynamics such as cerebral blood volume (CBV), cerebral blood flow (CBF), mean transit time (MTT) can be relatively measured and mapped [3].

However, quantification by the deconvolution method requires an arterial input function (AIF) obtained from the major cerebral arteries, and is thought to be highly susceptible to signal noise and technical error [4]. While evaluating perfusion parameters, algorithm is analyzing time steps images on a pixel-by-pixel basis. It is obvious that small noise can change pixel intensity

thus values of concentration time curve (CTC), which leads to artifacts in perfusion parameters maps.

Classic and widely used approach to reduce noise on perfusion images is filtering (smoothing or blurring). Often very large kernels are used, even 5×5 . These has a great effect on the quality of the image, which already is rather small (128×128 pixels). Such radical smoothing/blurring of the acquired data leads to losing a lot of information. Therefore we are proposing a technique, based on the interpolation, that is not affecting the image itself but is using values of pixels in the neighborhood of the one being analyzed to calculate a value of CTC in a point.

2 MRI Data Analysis

The perfusion study consists of series of images acquired in a short period of time after administration of the contrast agent. To perform such quick image acquisition, magnetic resonance is set to lower matrix (128×128) and interslice gap is increased. As a result we obtain several hundreds of images, divided among several dozens of time steps. For example 45 time steps, 10 slices each.

The physical process which give rise to a signal drop as a contrast agent passes through the tissue is attributed to a magnetic spoiling effect on the local MR resonance process. The signal is monitored over time with typical temporal resolution of the order of seconds (Fig. [1](#)). The bolus itself will pass entirely through the head in around ten seconds. The effect on signal is larger where interface between capillaries and body cells. If this barrier is impermeable to the contrast agent then this is the only source of signal loss and the effect is greatest for small capillaries for a given fixed volume of contrast. Signal loss in arteries and veins is due to direct mixing of blood and contrast agent. As the size of the capillaries in a given region of the tissue is unmeasurable all analysis methods assume a direct relationship between signal loss and total contrast volume within a voxel. The assumption of a linea relationship between relaxation rate and concentration of the contrast agent has been shown to be valid both by experiment and simulation for the blood volume fractions in the physiological and pathological range [\[4\]](#).

In classic approach, the first step of analysis is noise removal, by smoothing with a 5×5 uniform smoothing kernel [\[5\]](#). In our case this step is omitted and background removal is performed. This can be achieved by a simple thresholding to the certain value. By analyzing the looks of the general histogram of the MRI perfusion slice we can find two main peaks on in. First on the left, represents numerous low frequency background pixels. With increasing intensity of pixels (to the right of the diagram) number of pixels decreases reaching local minimum (S_0) and increasing again to reach another local maximum (S_B). In general case of MRI study, segmentation at the level S_0 would be sufficient, but in the case of the perfusion images we need a "deeper" segmentation. Therefore we have set the threshold value at $S_P = S_0 + \frac{(S_B - S_0)}{4}$.

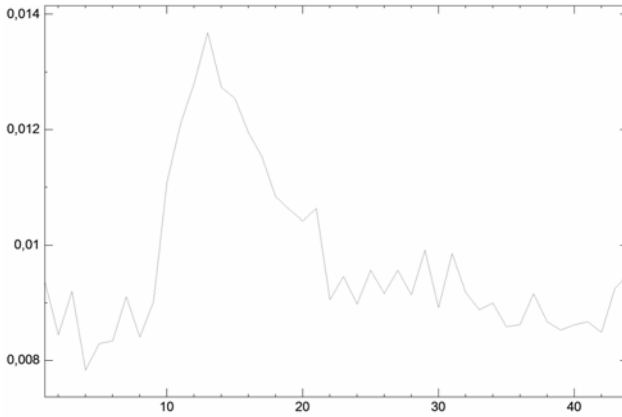


Fig. 1. Example of concentration time curve

AIF function was obtained from circular regions of interests (ROI) placed manually on the middle cerebral artery (MCA). To reduce the partial volume effect (PVE) we have searched the ROI for the maximal intensity value and chosen only pixels with intensity $> 99\%$ of this value. Mean concentration time curve was then calculated from these pixels and used as the AIF. We have used left MCA AIF to calculate the CBF, CBV and MTT of the left hemisphere and the right AIF MCA for the right hemisphere.

The signal intensities in the brain parenchyma were converted to transverse relaxation rates on the voxel-by-voxel basis. The signal time curves $I(t)$ were used to evaluate concentration time curve ($C(t)$) with following equation:

$$C(t) = -K * \ln \frac{I(t)}{I_0} \quad (1)$$

where: I_0 is the precontrast baseline signal, $I(t)$ is the signal intensity at time t after injection of contrast agent; $K = \frac{1}{TE}$; TE is the echo time. Next $C(t)$ was deconvoluted with AIF ($AIF(t)$) by using fast Fourier transformation to obtain $C_d(t)$ [6, 7] using equation:

$$C_d(t) = F^{-1}\left(\frac{F(C(t))}{F(AIF(t))}\right) \quad (2)$$

Following formula was used to calculate CBV:

$$CBV = \frac{k_h}{\rho} \frac{\int_0^{\infty} C(t) dt}{\int_0^{\infty} AIF(t) dt} \quad (3)$$

where k_h is a factor (0.73) correcting for the difference in hematocrit between the capillaries and large arterial vessels [8] and ρ is the density of the brain tissue ($1.04g/cm^3$).

CBF is:

$$CBF = \frac{k_h}{\rho} C_{max} \quad (4)$$

where C_{max} is the maximal value of $C_d(t)$.

MTT was calculated as the ratio:

$$MTT = \frac{CBV}{CBF} \quad (5)$$

These values were calculated for each slice acquired for the analyzed period of time (in our case each study had 10 slices per time step) on the pixel-by-pixel basis. These lead to making spatial maps of perfusion parameters. Unfortunately without noise removal maps contained some artifacts, that made it difficult to interpret the results of perfusion analysis. To reduce them we propose following method which has to be applied at the moment of sampling the signal time curve (1).

3 Interpolated Pixel Sampling

Lets assume that we want to evaluate function value in point $x \in R$ knowing its values in surrounding points. Function value I is:

$$I = \sum_k I(k) \cdot K(x - k) \quad (6)$$

where $K(x)$ is continuous function named the interpolation kernel. Summation is done for k points neighboring with point x . Without losing the generality and to simplify calculations we assume that the distance between consecutive k points is 1.

(7) is a general formula for calculation of discrete signal value in a real point. We are using symmetrical and separable interpolation kernels [9], therefore we can write formula for interpolation of the two dimensional signal as :

$$I(x, y) = \sum_k \sum_l I(k, l) \cdot K_{2D}(x - k, y - l) \quad (7)$$

where $K_{2D}(x, y) = K(x) \cdot K(y)$. In our deliberations we will understand $I(x, y)$ as an intensity of voxel of perfusion image in point (x, y) .

As mentioned earlier, image sampling for the calculation of concentration time curve (1) is done on a pixel-by-pixel basis. To ensure that sampled pixel P value is not affected by the noise, we analyze its neighborhood in the manner shown on the Fig. 2. We consider pixels aligned on horizontal, vertical and two diagonal lines passing through analyzed pixel on the plane of the slice. Note: because of the great distance between slices in the series, it is not possible to consider voxels from below and above slices.

We assume that the P is not know and interpolate it's value in four, one dimensional directions. In case of the diagonal neighborhood we assume distance, between consecutive pixels, to be equal $\sqrt{2}$. As the result we get four

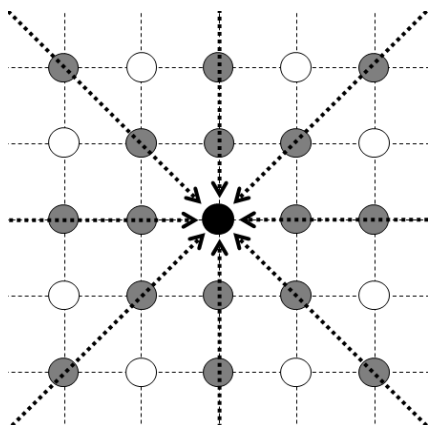


Fig. 2. Interpolated pixel sampling

values of intensities. After excluding minimum and maximum value, average from two remaining values is calculated. This should minimize influence of possible noise in surrounding pixels. Next this average is compared with actual signal intensity in point P . If the difference is larger than 10%, interpolated value is chosen as the value of point P , in other case original signal value is taken.

In case of borders of the brain, that is situation when some consecutive pixels on the analyzed lines are equal zero, method is not used, as it could bring more harm than benefits. Especially when information from the edge of the brain rarely presents data of diagnostic importance. The conditions for applying above method are:

- interpolation along the 1D line can be performed if at least $\frac{2}{3}$ of significant (belonging to interpolation kernel) points are > 0
- interpolation along at least 1 line is possible.

4 Tested Methods

Because the portfolio of existing interpolation kernels is very wide we have chosen only several methods for analysis. Table 1 shows methods we have tested. We've used simple and more advance methods.

These are most widely used interpolation methods. They are evaluating function value using different number of neighboring pixels. The simplest use only two neighbors, but more complex (for example *sinc* family) can use larger number of points. Because *sinc* function, by definition, is infinite it must be multiplied by specific window. Research shows that the best results are obtained by odd sized kernels [17].

Table 1. Tested interpolation methods

Method name	Formula	Ref.
Linear interpolation	$K_{lin}(x) = \begin{cases} 1 - x , & 0 \leq x < 1; \\ 0, & \text{else.} \end{cases}$	[10]
Cosine interpolation	$K_{cos}(x) = \begin{cases} \frac{(1 - (1 - \cos(x \cdot \pi)))}{2}, & 0 \leq x < 1; \\ 0, & \text{else.} \end{cases}$	[11]
Cubic interpolation	$K_{cub}(x) = \begin{cases} (\alpha + 2) x ^3 + (\alpha + 3) x ^2 + 1, & 0 \leq x < 1; \\ \alpha x ^3 - 5\alpha x ^2 + 8\alpha x - 4\alpha, & 1 \leq x < 2; \\ 0, & \text{else.} \end{cases}$	[12]
Spline Catmull-Rom interpolation	$I(x) = 0.5 \cdot (1.0 \ x \ x^2 \ x^3) * \begin{pmatrix} 0 & 2 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 2 & -5 & 4 & -1 \\ -1 & 3 & -3 & 1 \end{pmatrix} * \begin{pmatrix} P_0 \\ P_1 \\ P_2 \\ P_3 \end{pmatrix}$	[13]
Sinc interpolation	$K_{sinc}(x) = \frac{\sin(\pi \cdot x)}{\pi \cdot x}$ $K_{N_{sinc}}(x) = \begin{cases} (K_{sinc}(x) \cdot w(x)), & 0 \leq x < \frac{N}{2}; \\ 0, & \text{else.} \end{cases}$	
Welch window	$w_{welch}(x, r) = \begin{cases} 1 - (\frac{x}{r})^2, & x < r; \\ 0, & \text{else.} \end{cases}$	[14]
Hann-Hamming window	$w_{hh}(x, r, \alpha) = \begin{cases} \alpha + (1 - \alpha) \cdot \cos(\pi \cdot \frac{x}{r}), & x < r; \\ 0, & \text{else.} \end{cases}$	[15]
Lanczos window	$w_{lanczos}(x, r) = \begin{cases} \sin(\pi \cdot \frac{x}{r}) / (\pi \cdot \frac{x}{r}), & x < r; \\ 0, & \text{else.} \end{cases}$	[16]

5 Tests Results

To objectively test our method we have created a set of ideal slices without any noise distortion. We were basing on the data from one of the clinics we are cooperating with and manually remove each abnormal pixel, creating almost clear images. Imaging protocol used consisted of perfusion study [echo planar imaging (EPI), TR/TE: 1940/80 msec, 10 slices with a 5mm thickness and 10mm interslice gap, matrix: 128×128]. For this ideal data set we have calculated all hemodynamics parameters and prepared map of CBV, CBF and MTT values. Next we have introduced artificial noise (of different strength) into the test study and tried different methods of noise influence reduction. We have used noise grains one and two pixels in diameter big, that were placed randomly with probability 5%, 10% and 20%. Noise pixels had maximal acceptable value. Results were then compared with ones obtained for the ideal study. Mean distance between values of each parameter (for each corresponding pixel) was calculated and used as the criterium for the quality assessment. Example results for these tests are show in the Table 2.

Table 2. Results of test - average distance between values of CBV (normalized)

method name	kernel size	grain $\phi 1$			grain $\phi 2$		
		5%	10%	20%	5%	10%	20%
Linear	2	0,0131	0,0231	0,0412	0,1022	0,1632	0,2843
Cosine	2	0,0113	0,0201	0,0356	0,1054	0,1778	0,2654
Cubic	4	0,0094	0,0121	0,0285	0,0687	0,1238	0,1843
Spline	4	0,0095	0,0163	0,0232	0,0621	0,1183	0,1743
Sinc Welch	5	0,0043	0,0082	0,0143	0,0354	0,0568	0,0872
Sinc Hann-Hamming	5	0,0035	0,0073	0,0112	0,0293	0,0432	0,0712
Sinc Lanczos	5	0,0038	0,0076	0,0121	0,0295	0,0456	0,0781
Filtering	3	0,0211	0,0356	0,0543	0,0844	0,1424	0,2172
	5	0,0308	0,0451	0,0811	0,1232	0,1804	0,3244
no noise reduction		0,0413	0,0803	0,1514	0,1651	0,3223	0,6050

After analyzing average distances between values of CBF, CBV and MTT, together with standard deviations, we have come to conclusion that interpolation methods are very effective in reducing noise effect, especially when compared with classic filtering approach. Lack of noise reduction leads to visible artifacts in perfusion parameters maps, which in turn can lead to misinterpretation. Therefore, in fact, some action must be done to minimize these artifacts. Traditional approach allows to do this, but has degrading influence on the image quality and in our opinion should be avoided, especially when proposed interpolation sampling produces very good results. Even such simple method as linear interpolation produces better results than mentioned filtering. Small kernel interpolation functions prove to be less effective when dealing with larger noise grains, but were very good for little, single pixel noise. Most universal, for small and larger noise grains, are very effective *Sinc* family methods, that are using 5-pixel large windows. Differences between *Sinc* windows are minimal, but Hann-Hamming window gives the best results. This method is unfortunately the slowest one, therefore Welch window should be chosen, as it produces similar results but executes 60% quicker.

6 Conclusions

In this article we have presented new approach to minimalization of disadvantageous noise influence on the evaluation of perfusion hemodynamics parameters. We have compared our technique with the methods classically used and showed that, in fact, using interpolation sampling gives better results than simple filtering.

References

1. Edelman, R.R., Siewert, B., Darby, D.G., Thagaraj, V., Nobre, A.C., Mesulam, M.M.: Qualitative Mapping of cerebral blood-flow and functional localization with echo-planar MR imaging and signal targeting with alternating radio-frequency. *Radiology* 192, 513–520 (1994)
2. Mihara, F., Kuwabara, Y., Tanaka, A., et al.: Reliability of mean transit time obtained using perfusion weighted MR imaging: comparison with positron emission tomography. *Magn. Reson Imaging* 20, 33–39 (2003)
3. Restrepo, L., Wityk, R.J., Grega, M., et al.: Diffusion and perfusion-weighted magnetic resonance imaging of the brain before and after coronary artery bypass grafting surgery. In: *Stroke* 2002, vol. 33, pp. 2909–2915 (2002)
4. Boxerman, J.L., Rosen, B.R., Weisskoff, R.M.: Signal-to-noise analysis of cerebral blood volume maps from dynamic NMR imaging studies. *J. Magn. Reson. Imag.* 7, 528–537 (1997)
5. Kaneko, K., Kuwabara, Y., Mihara, F., et al.: Validation of the CBF, CBV, and MTT values by perfusion MRI in chronic occlusive cerebrovascular disease: a comparison with OPET. *Academic Radiology* 11(5), 489–497 (2004)
6. Wirestam, R., Anderson, L., Ostergaard, L., et al.: Assessment of regional blood flow by dynamic susceptibility contrast MRI using different deconvolution techniques. *Magn. Reson. Med.* 43, 691–700 (2000)
7. Ostergaard, L., Weisskoff, R.M., Chester, D.A., et al.: High resolution measurement of cerebral blood flow using intravascular tracer bolus passages Part I: Mathematical approach and statistical analysis. *Magn. Reson. Med.* 36, 715–736 (1996)
8. Berczki, D., Wei, L., Otsuka, T., et al.: Hypercapnia slightly rises blood volume and sizably elevates flow velocity in brain microvessels. *M. J. Physiol.* 264, H1360–H1369 (1994)
9. Lehmann, T.M., Gönner, C., Spitzer, K.: Survey: Interpolation Methods in Medical Image Processing. *IEEE Transactions on medical imaging* 18(11) (1999)
10. Parker, J.A., Kenyon, R.V., Troxel, D.E.: Comparison of interpolating methods for image resampling. *IEEE Trans. Med. Imag.* MI-2, 31–39 (1983)
11. Beauchamp, K.G.: *Signal Processing*. George & Allen Unwin Ltd (1973)
12. Danielsson, P.E., Hammerin, M.: High accuracy rotation of images. *Computer Vision, Graphics and Image Processing* 54(4), 340–344 (1992)
13. Catmull, E., Rom, R.: A class of local interpolating splines. In: Barnhill, R.E., Riesenfeld, R.F. (eds.) *Computer Aided Geometric Design*, pp. 317–326. Academic Press, New York (1974)
14. Meijering, E.H.W.: Spline interpolation in medical imaging: comparison with other convolution-based approaches. In: *Signal Processing X: Theories and Applications - Proceedings of EUSIPCO 2000*, vol. IV, pp. 1989–1996 (2000)
15. Wolberg, G.: *Digital Image Warping*. IEEE Computer Society Press, Los Alamitos (1990)
16. Lanczos, C.: *Discourse on Fourier Series*. Oliver & Boyd, London (1966)
17. Thacker, N.A., Jackson, A., Moriarty, D., Vokurka, B.: Renormalised Sinc Interpolation, University of Manchester, Tina Memo No. 1999–005 (1999)

Interpretation of the Sequences of Flow Magnetocardiographical Images Based on Flow of Electrical Impulses through Human Heart

Kamila Baron-Pałucka

AGH University of Science and Technology,
Institute of Automatics, 30 Mickiewicza Ave, 30-059 Krakow, Poland
e-mail: kamila.baron@agh.edu.pl

Summary. Magnetocardiography is a promising diagnostic technique that is a magnetic equivalent of electrocardiography. This article presents a novel approach to interpretation of the sequences of magnetocardiographical images that could support diagnostics of cardiac infarctions. Algorithm presented in this paper was used to create a system that classifies to or excludes patients from cardiac infarction risk group.

1 Introduction

Magnetocardiographic examination (MCG) registers intensity of the magnetic field generated during cardiac electrical activity and can be considered magnetic equivalent of electrocardiographic examination (ECG), but MCG is more sensitive in detection of magnetic field generated by tangential currents than ECG is in detection of electrical field generated by the same currents [6]. Furthermore, vortex currents generate magnetic field detected by MCG, but they are not the source of any electrical field that could be detected by ECG [2]. Those features make MCG more suitable for detection of the diseases that are caused by alteration of the direction of the electrical currents flowing through the heart. Since in healthy heart main direction of the activation wavefront is radial, from endocardium to epicardium, then MCG enables more precise detection of the ischemic changes in direction of depolarization and repolarization than ECG does.

This article presents a novel algorithm of patient's MCG data evaluation. Instead of using well known diagnostic parameters described in literature - those that use similarity of MCG and ECG time runs, as well as those computed based on latter described magnetic field maps, like the alteration in the direction of the magnetic induction vector - the alternative approach is proposed. Decision about patient's classification to heart disease risk group is made based on latter described trajectories that are related to the flow of electrical impulses in the heart.

2 MCG Data Formats

The values of the intensity of the magnetic field that are registered in course of a magnetocardiographic examination can be presented in three main formats.

2.1 Time Runs

The values of the intensity of the magnetic field can be registered in form of time runs - each measuring point over patient torso is associated with one time run. Morphological features of MCG and ECG time runs are lot alike - on MCG time runs there are parts similar to P wave, QRS complex and T and U waves from ECG time run - there is also timing correlation between those elements. The significant difference is the fact, that in measuring points placed over lower thorax, in proximity of the midsternal plane, time runs with normal orientation of R and T waves are registered, whereas in measuring points placed over upper left thorax, time runs have reversed orientation.

2.2 Magnetic Field Maps (MF Maps)

By virtue of similarities between ECG and MCG time runs, MCG data in time runs format can be interpreted basing on rules created for ECG data [3], but the alternative approach is to convert it to map format that could be subject of further analysis. MF maps are created by computing spatial distribution of the measured magnetic field intensity values for each channel (measuring point) in the same time point, as well as of the interpolated values for points placed between sensors - the details are beyond scope of this paper, but can be found in [4]. As a result of this operation, contour map or equivalent map with artificial coloring scheme is developed. Fig. 1 demonstrates correlation between electrical vectors of the heart in particular phases of cardiac cycle, and MF maps.

The visualization method that is frequently used for presentation of data collected in course of the MCG examination is a development of a sequence of

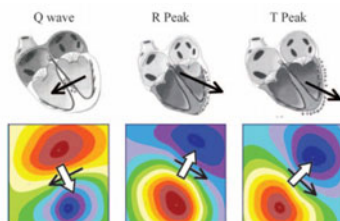


Fig. 1. Correlation between electrical vectors of the heart in particular phases of cardiac cycle and MF maps. Electrical vector is depicted with black arrow whereas vector of the magnetic field is depicted with white arrow. In all phases, as expected, both vectors are mutually orthogonal [2].

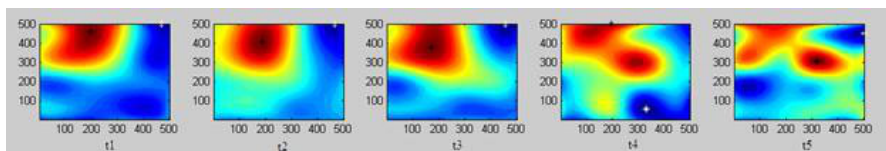


Fig. 2. First five maps from sample MF maps sequence that consists of 1000 maps spanning through one complete heart cycle

the maps computed for consecutive moments of the cardiac cycle, as presented in Fig. 2. Such visualization, often in form of animation, allows observer to capture temporal-spatial dynamics of the alterations on MF maps related to heart functioning.

2.3 Pseudo Current Density Maps (PCD Maps)

PCD (Pseudo Current Density) maps were introduced to enable such representation of the magnetic field intensity values, that would reflect the source of the measured values - the distribution of the density of the heart’s currents [5]. The conversion to that format is accomplished by applying to B_z component of the magnetic induction vector the HC (Hosaka-Cohen) transformation, that allows to obtain the value of so-called pseudo current density.

The application of PCD maps in visualization of MCG data, results in maps where localization of the point with the largest signal amplitude is equivalent to the localization of the electrical dipole of the heart, and orientation of this point’s pseudo current density vector is in accordance with orientation of electrical dipole of the heart. Therefore, PCD map is intuitive for doctor’s interpretation, since it reflects which areas of the heart are active at the moment.

Transformation from MF to PCD format is described by following equation:

$$\vec{c} = \frac{\partial B_z}{\partial y} \cdot \vec{e}_x - \frac{\partial B_z}{\partial x} \cdot \vec{e}_y \tag{1}$$

where: \vec{c} - pseudo current density B_z - component of the magnetic induction vector (registered in course of MCG examination) \vec{e}_x, \vec{e}_y - versors of the coordinate system

Fig. 3 shows PCD maps that were created from the MF maps presented in Fig. 2.

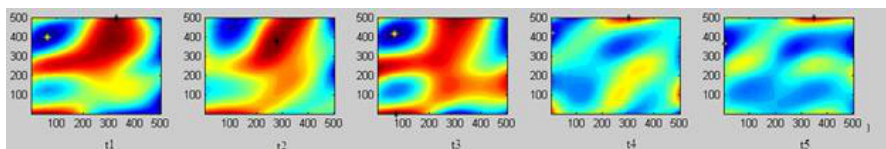


Fig. 3. PCD maps created for MF maps from Fig. 2

3 Novel Approach to Patient's MCG Data Evaluation

The most common way to analyze magnetocardiographic data involves the usage of MCG time runs and computing morphological parameters that base on similarities between MCG and ECG signals. The other way is to attempt to directly utilize MF map sequences, and make diagnoses based on parameters like [1, 3]:

- α angle - tilt angle of the vector that joins points of the largest (+) and the smallest (-) amplitude on MF map
- rotation of α angle in 30ms
- variation of the distance between (+) and (-) points in 30ms
- variation of the ratio between the field strength in points (+) and (-) in 30ms

PCD maps help visualize the flow of the electrical currents that generate heart's magnetic field. Patients after stroke usually suffer from necrotizing damages in heart's muscle - those changes can alter the standard paths through which electrical signals should flow in order to cause contraction and relaxation of atria and chambers. PCD maps can potentially show those alterations and thus help to assess whether patient should be classified to heart disease risk group. This observation was an impulse for an attempt of creating a method, that would assess MCG data from the perspective related to the flow of electrical impulses in the heart.

In method proposed by author such first attempt is made, and patient's classification is obtained in following steps:

1. Sequences of MF and PCD maps are created that span over one complete cardiac cycle of the patient
2. On each map of the sequence two points are identified - the maximum and the minimum
3. The coordinates of the identified points are sequentially marked on four separate maps in result constituting four types of trajectories:
 - trajectory of the point of the maximal value on sequence of MF maps
 - trajectory of the point of the minimal value on sequence of MF maps
 - trajectory of the point of the maximal value on sequence of PCD maps
 - trajectory of the point of the minimal value on sequence of PCD mapsCurrently all four types of trajectories for one patient in question are used, in order to determine how effective are classifications based on certain trajectory type.
4. Classification of the patient to heart disease risk group is based on comparing his trajectories with trajectories of patients in two test groups - test patients diagnosed with heart disease and healthy test patients without history of heart diseases
5. The comparison is made with the use of Hausdorff metric - patient is classified to the group for which minimal Hausdorff distance between his trajectory and test patient trajectory was obtained

4 Experiment and Discussion

Following experiment was performed to validate that classification based on trajectories comparison can potentially be accurate method.

4.1 Experimental Database

Available research material in form of MCG data records was acquired in course of diagnostic examinations performed in The John Paul II Hospital in Krakow, with the use of CMI 2409 magnetocardiograph, CardioMag Imaging Inc. Examinations were performed on patients already diagnosed with other techniques and remaining under medical consultancy. After elimination of incomplete or incorrectly registered records, database contains results of 276 examinations that include information whether patient was (62) or wasn't (214) previously diagnosed with cardiac infarction.

4.2 Test Groups Construction

Two test groups were created:

- "Stroke" test group - 10 patients randomly chosen from complete set of patients whose records were marked as diagnosed with cardiac infarction
- "Healthy" test group - 10 patients randomly chosen from complete set of patients whose records were marked as not diagnosed with cardiac infarction

4.3 Parameters of the Experiment

Interpolation step

Data set collected for each patient with the use of CMI 2409 magnetocardiograph consists of data from 36 channels (6×6) that were registered in 1000 time points spanning through one complete cardiac cycle. In order to create from 36 original data points a complete MF map, preprocessing phase included interpolation. The considerations concerning the influence of different interpolation methods on maps quality and computational effort are beyond the scope of this paper, but can be found in [4]. Nevertheless, data generated with different interpolation steps was compared in course of the experiment to assess its influence on obtained results. Five different interpolation step values were explored, 1, 0.5, 0.2, 0.1 and 0.01.

Part of heart cycle

According to literature [3], diagnostic parameters should be calculated for the time frames that correspond to the type of the disease being diagnosed - for example in case of ischemic heart diseases only ventricular repolarization phase

should be taken into consideration. To confirm this, in course of experiment the complete heart cycle was divided into five equal parts and trajectories created for each of them were classified to enable comparison of classification results. This means that trajectories were created for following intervals - samples 1:200, 201:400, 401:600, 601:800 and 801:1000.

Sampling

Since trajectories were computed for intervals of 200 samples length, that would mean that if all samples were to be taken into consideration Hausdorff metric would have to compare shapes described by 200 points each. In course of experiment the influence of different sampling density on final results was assessed - from original 200 points following subsets were created - trajectory of 100 samples length (every 2nd original sample), trajectory of 50 samples length (every 4th original sample) and trajectory of 20 samples length (every 10th original sample).

4.4 Results Summary

Tables 1, 2, 3 and 4 summarize the best classification results obtained in course of the experiment.

4.5 Results Discussion

The results obtained in course of the experiment allow to draw following conclusions:

1. The best classification results are obtained in last part of heart cycle (samples 801-1000) that roughly corresponds to ventricular repolarization
2. The best classification results are obtained with the usage of trajectories created based on PCD maps sequence

Table 1. The best classification results, from interpolation step perspective, in group of patients previously diagnosed with cardiac infarction ("Stroke") and in group of patients without history of cardiac infarction ("Healthy").

Interpolation step	Interval start	Interval stop	Sampling (every x-th sample)	Correctly classified "Stroke" patients	Correctly classified "Healthy" patients	Trajectory type
1	801	1000	2	58%	46%	MIN PCD
0.5	801	1000	2	73%	44%	MIN PCD
0.2	1	200	10	63%	37%	MIN PCD
0.1	801	1000	2	63%	39%	MIN PCD
0.01	601	800	4	60%	49%	MAX MF

Table 2. The best classification results, from heart cycle part perspective, in group of patients previously diagnosed with cardiac infarction ("Stroke") and in group of patients without history of cardiac infarction ("Healthy")

Interval start	Interval stop	Interpolation step	Sampling (every x-th sample)	Correctly classified "Stroke" patients	Correctly classified "Healthy" patients	Trajectory type
1	200	0.2	10	63%	37%	MIN PCD
201	400	0.5	10	58%	49%	MAX MF
401	600	0.1	10	50%	53%	MAX MF
601	800	0.01	4	60%	49%	MAX MF
801	1000	0.5	2	73%	44%	MIN PCD

Table 3. The best classification results, from sampling perspective, in group of patients previously diagnosed with cardiac infarction ("Stroke") and in group of patients without history of cardiac infarction ("Healthy")

Sampling (every x-th sample)	Interpolation step	Interval start	Interval stop	Correctly classified "Stroke" patients	Correctly classified "Healthy" patients	Trajectory type
2	0.5	801	1000	73%	44%	MIN PCD
4	0.5	801	1000	71%	43%	MIN PCD
10	0.2	1	200	63%	37%	MIN PCD

Table 4. The best classification results, from trajectory type perspective, in group of patients previously diagnosed with cardiac infarction ("Stroke") and in group of patients without history of cardiac infarction ("Healthy")

Trajectory type	Interpolation step	Interval start	Interval stop	Sampling (every x-th sample)	Correctly classified "Stroke" patients	Correctly classified "Healthy" patients
MAX MF	0.2	801	1000	10	63%	32%
MIN MF	0.01	801	1000	4	58%	57%
MAX PCD	0.01	601	800	4	58%	39%
MIN PCD	0.5	801	1000	2	73%	44%

3. Experiment results suggest that patients without history of cardiac infarction were underrepresented in test group - 10 out of 214 in comparison to 10 out of 62 for patients previously diagnosed with cardiac infarction. This was a probable cause of classification results for that group being worse than expected

4. Experiment results suggest that the optimal value of interpolation step might be 0.5, but it needs further analysis to confirm it

5 Conclusions

The novel approach to interpretation of MCG maps sequences seems to be a promising start towards utilization of information about electrical impulses flow in cardiac infarction diagnosis. The experiment revealed following areas that will need to be addressed in author's future research:

- The size of new test groups should be correlated with the size of represented populations
- More precise identification of ventricular repolarization phase should be performed to assess if that could improve classification results
- Methods alternative to Hausdorff metric could be used to evaluate trajectories similarity

References

1. Fenici, R., Brisinda, D., Meloni, A.: Clinical application of magnetocardiography. In: Expert Review of Molecular Diagnostics, pp. 291–305 (2005)
2. Steinberg, B., Roguin, A., Watkins III, S.: Magnetocardiogram recordings in a nonshielded environment - reproducibility and ischemia detection. *Annals of Noninvasive Electrocardiology* 10(2), 152–160 (2005)
3. Chen, J., Thomson, P., Nolan, V.: Age and sex dependent variations in the normal magnetocardiogram compared with changes associated with ischemia. *Annals of Biomedical Engineering* 32(8), 1088–1099 (2004)
4. Baron-Pałucka, K.: MCG data preprocessing. *Bio-Algorithms and Med-Systems* (2010) (in press)
5. Haberkorn, W., Steinhoff, U., Burghoff, M.: Pseudocurrent density maps of electrophysiological heart, nerve or brain function and their physical basis. *BioMagnetic Research and Technology* 4(5) (2006)
6. Baron-Pałucka, K.: Magnetocardiography - measurement conduction, data formats and diagnostic applications. *Bio-Algorithms and Med-Systems* 4(7), 25–34 (2008)

Automatic Left Ventricle Segmentation in T2 Weighted CMR Images

Kurhairy Abdul Kadir¹, A. Payne², John J. Soraghan¹, and C. Berry²

¹ Centre for excellence in Signal and Image Processing/Dept of Electronic and Electrical, University of Strathclyde

e-mail: kushsairy@eee.strath.ac.uk

² Golden Jubilee National Hospital, Glasgow

Tel.: +44 141 548 2205; Fax:+44 141 552 2487

Summary. An automatic left ventricle (LV) segmentation method for T2 weighted Cardiac Magnetic Resonance (CMR) image is presented. The method takes multi-slice T2 weighted CMR images from the basal to the apex of the heart. Inter-slice and intra-slice fuzzy reasoning is used to guide the centre point detection for each slice. Morphological filtering is used in the reconstruction to homogenise the blood pool region. Then radial search Fuzzy Multiscale Edge Detection (FMED) is used to segment the endocardium and the epicardium of the LV. Evaluation of the method is performed on 6 patient with approximately 42 slices of real T2 weighted MRI data. The quantitative result of the automatic method compared to those obtained from manual segmentation by a skilled clinician are very encouraging, with correlation scores of 96.2% correlation for the endocardium and 85.7% correlation for the epicardium.

1 Introduction

The level of injury after myocardial infarction and the distinction between viable and dead myocardial tissue are important for cardiac treatment decision making. Oedema is fluid retention in myocardial tissue due to damage tissue causing swelling in the affected area. Oedema appears as bright areas compared to the normal myocardial tissue. The importance of studying oedema in MI is that it provides a means of differentiating acute and chronic MI [1], thus T2 weighted cardiac magnetic resonance (CMR) images provide mean to visualise normal area and oedema area on myocardial wall after myocardial infarction (MI).

In order to determine the location and quantification of oedema in myocardial tissue, the first step is to segment the endocardium and epicardium of left ventricles on every short axis (SA). This process is tedious and time consuming when carried out manually. The challenges in automatically segmenting LV include dealing with the presence of papillary muscles in the

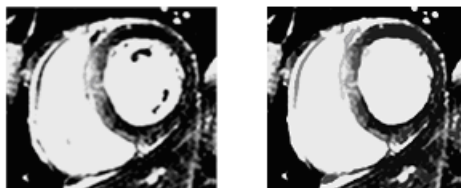


Fig. 1. (left) Original image and filtered image (right)

endocardium wall, weak edges around the epicardium of the LV, and the fact that the presence of oedema in the myocardial tissues has higher signal intensity compared to normal myocardial tissue and the precise location of the oedema cannot be predicted further complicated the segmentation process.

Several studies involving viability assessment of myocardial tissue used either manual [2], [3] or semi automatic technique to segment the border of the left ventricle (LV). Stalidis et al. [4] applies contour deformable model to segment CMR image. The technique requires the user to select several points from each class for the calculation of the mean and the standard deviation. Ciofalo et al [5] applies a fully automatic segmentation using combinations of technique such a geometrical template and shape prior model to segment the LV, but the work is applied to late-enhancement CMR image.

In this paper we present a new method to segment edocardium and epicardium of LV using data driven fuzzy reasoning in a multiscale environment. The paper is organised as follows. Section 2 describes the various sub-components of the new multi-slice LV segmentation process. Section 3 presents results of the using the method on real acquired T2 weighted CMR data. The same data sets were subsequently analysed by a trained clinician and the results compare favourably to those obtained from the automatic method. Section 4 concludes the paper.

2 Automatic Left Ventricle Segmentation System

Each slice from the CMR image is processed using fuzzy based automatic LV segmentation system comprising of *Pre Processing, Multislice Fuzzy Based Centre Point Detection, and Multislice Fuzzy Based Edge LV Edge Detection.*

2.1 Pre-processing and Centre Point Detection

Geodesic morphological dilation [6] is used to pre-process the image to create homogenous region within the blood pool of the LV. Fig. 1 illustrated an MRI signal before and after the reconstruction, where the dark area in the blood pool cavity is seen to be converted to the same intensity level to the blood surrounding it. Centre point detection is an important step in segmenting the LV. In our implementation we incorporate the knowledge of the LV centre

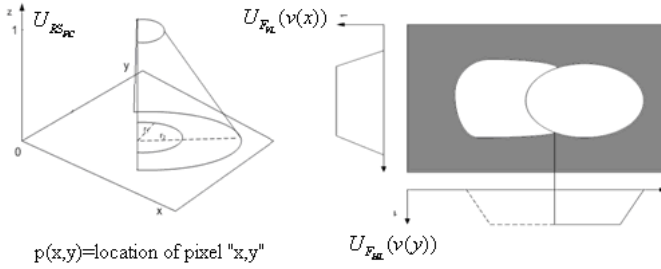


Fig. 2. (left) centre proximity, Intensity profile of LV (right)

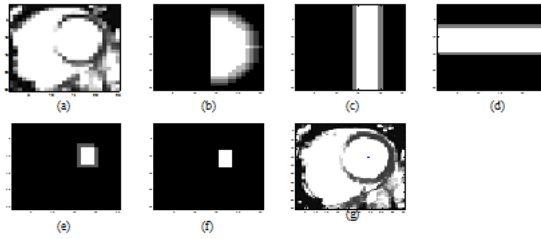


Fig. 3. (a) reduced resolution image, fuzzy sub-sets (b) RS_{FC} , (c) F_{VL} , (d) F_{HL} , (e) $F_{intersect}$, (f) $F_{\alpha cut}$, (g) C_{point}

point in CMR image in a fuzzy way to classify a few number of pixels points as the possible candidate for the centre point.

- Right sided fuzzy centre (RS_{FC})
- Fuzzy vertical line (F_{VL})
- Fuzzy horizontal line (F_{HL})

Using reduced resolution CMR image [7] as shown in Fig. 3a, three corresponding fuzzy sub sets which are the same sizes with the reduce resolution image can be represented as:

$$RS_{FC} = \{(x, y), \mu_{RS_{FC}}(p(x, y))\}, (x, y) \in \{image\ pixel}\} \quad (1)$$

$$F_{VL} = \{x, u_{F_{VL}}(v(x))\}, x \in \{index\ of\ vertical\ scan}\} \quad (2)$$

$$F_{HL} = \{y, u_{F_{HL}}(v(y))\}, y \in \{index\ of\ horizontal\ scan}\} \quad (3)$$

where $p(x, y)$ is the position of pixel (x,y), $v(x)$ is the position of point x, and $v(y)$ is the position of point y. As well as that, $\mu_{RS_{FC}}(p(x, y))$, $u_{F_{VL}}(v(x))$, and $u_{F_{HL}}(v(y))$ are the degree of membership elements of image pixel, vertical scan and horizontal scan respectively. Fig. 2a illustrates the degree of

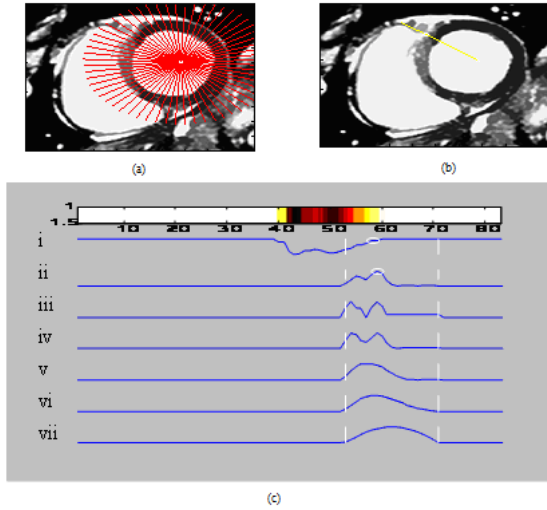


Fig. 4. (a) radial lines, (b) edge point, (c)(i) gradient profile, (ii) maximum value, (iii) to (vii) fuzzy subsets on each scales

membership element of RS_{FC} where there is no element to the left of the image plane and membership value decreases as the pixel moves away from the centre. **2b** illustrates intensity profile for vertical scanning (F_{VL}) along x axis and horizontal scanning (F_{HL}) along y axis. The fuzzy subsets are illustrated in Fig. **3b**, **3c** and **3d**. Once the knowledge has been represented in fuzzy way, the rules are then fed into the fuzzy inference engine by means of fuzzy 'min' operator to produce pixel candidates for the centre point

$$F_{intersect} = \min(RS_{FC} \cap F_{VL} \cap F_{HL}) \quad (4)$$

Finally the candidates is found by applying fuzzy alpha cut ($F_{\alpha cut}$) where α is set empirically at 0.85 to further eliminate the pixel candidates. Fig. **3e** and Fig. **3f** illustrate the $F_{intersect}$, and $F_{\alpha cut}$ respectively. The final pixel point is determined by determining the geometric centre of the non-zero $F_{\alpha cut}$ as illustrated in Fig. **3g**.

2.2 Left Ventricle Segmentation

A fuzzy multiresolution edge detection (FMED) is used as a basis for our LV endocardium and epicardium segmentation algorithm. Fig. **4a** shows the radial line emanating from the centre point. In Fig. **4b** shows an edge location on a single line after processed by FMED. The process of finding the edge is shown in Fig. **4c**. Fig. **4c** (iii) to (iv) are membership values that appear on each scales. The fuzzy memberships are then combined and the final edge point is determined by finding the maximum as shown in Fig. **4c** (ii). This

point correspond to the edge as shown in Fig. 4 (i) on the gradient profile of the signal.

Endocardium and Epicardium segmentation

The processes of segmenting the endocardium include

1. Apply FMED to detect edges of the endocardium
2. Once all the endocardium edges have been found, a uniform cubic B-spline is used to smooth the boundary.
3. The edges on the first slice of the endocardium are then set as reference edges for the next slice. This interslice processing ensures continuity between slices.
4. The process from (2) to (3) is repeated up to the final slice of the LV.

Segmentation of epicardium for T2 MRI signal poses two major challenges (i) the weak edges between the edge of myocardium and the surrounding organs and (ii) the oedema tissues having higher intensity compared to the normal tissues. The steps used to segment the epicardium are:

1. A new radial line length is estimated, this is done by estimating the thickness of the interventricular septum between the LV and RV which provide a reliable estimate of the thickness for the rest of the myocardium.

$$r_{new} = r_{old} + r_{septum} \quad (5)$$

2. Segmentation process for the epicardium starts in the interventricular septum area. This is due to the edge information is stronger around this area. The averaging of these edge points ($edge_{avg}$) is used as an edge point reference for regions outside the RV as follows. Using this average, any new edge is tested as follows:

if $k \text{ edge}_{avg} \Rightarrow \text{edge}_{point} \Leftarrow k \text{ edge}_{avg}$
then $\text{edge}_{point} = \text{edge}_{point}$
else $\text{edge}_{point} = \text{edge}_{avg}$
end

Parameter k was determined by experiment.

3. A uniform cubic B-spline approximation is used to smooth the boundary.
4. The process is repeated up to the final slice of the LV.

Papillary muscle removal

Papillary muscles are significant around the mid-ventricle slice, however in order to quantify oedema it is important that these are removed from the

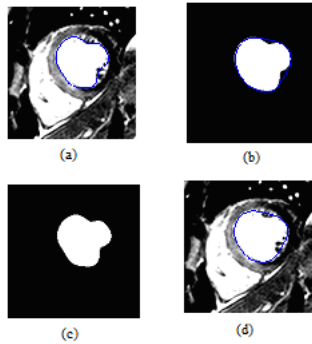


Fig. 5. (a) result of FMED with papillary muscle, (b) binary mask, (c) convex hull of the LV blood pool, (d) result with papillary muscle removed

endocardium analysis. Fig. 5 shows the process of removing the papillary muscles. These include:

- (i) The creation of a binary mask of the endocardium Fig. 5b
- (ii) Compute the convex hull of the LV blood pool Fig. 5c
- (iii) Superimpose the segmented result onto the original endocardium Fig. 5d.

3 Experimental Result

This section presents experimental result of the system. CMR images from 6 adult patient who recently experienced myocardial infarction were acquired using a 1.5 Tesla MRI machine (scanner: Siemens, Magneto Avanto 1.5T MRI system), for each patient 9 to 12 images are acquired. All short axis image were used except for the extreme basal and apical slices.

The results presented in Fig. 6 shows that those from the automatic system compared favourably to those from a manual segmentation by a skilled clinician. In Fig. 6a the automatic segmentation of the endocardium is consistent with the manual segmentation but for the epicardium the automatic segmentation overestimates the epicardium, especially in the area where edge information is very poor. In the mid ventricle area the automatic segmentation underestimate the endocardium area due to heavy present of papillary muscles as shown in Fig. 6b. In the apex slice the automatic segmentation is consistent with the manual segmentation.

Fig. 7 compares the area bounded by the endocardium boundary and the area bounded by the epicardium boundary for the automatic method and the manual segmentation processes. A total of 6 real T2 weighted data sets have been used. Good correlation is noted in both cases. For the area bounded by the endocardium 96.2% correlation is achieved and for the area bounded by the epicardium 85.7% correlation is achieved. For epicardium, the results obtained for the segmentation are not as good as compared to the results

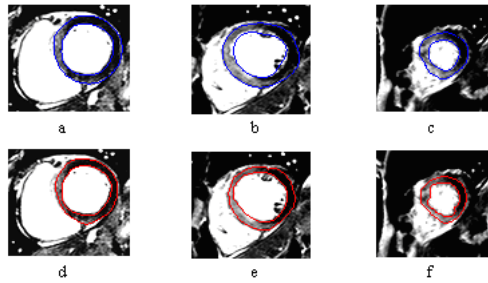


Fig. 6. Comparison between (a) to (c) automatic segmentation vs. manual segmentation (d) to (e), for basal level (a) & (d), median level (b) & (e), and apical level (c) & (f)

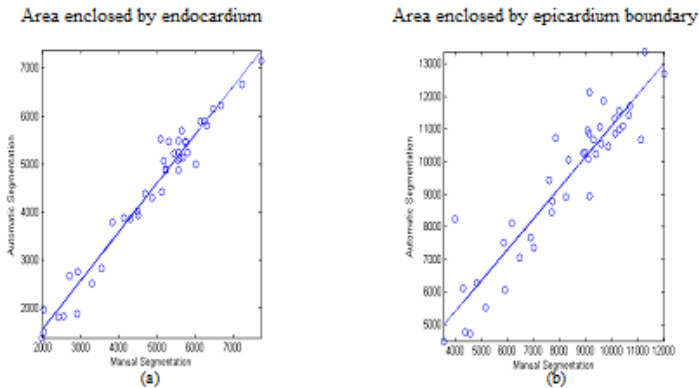


Fig. 7. Linear regression analysis with correlation score of 96.2% for endocardium and (b) 85.7% for epicardium

obtained for endocardium segmentation. There are few reasons for this (i) the contrast between the fat surrounding the heart and the lung is much stronger than the desired edge between the myocardium (ii) in certain area there is almost no edge information at all and (iii) the effect of high contrast of oedema area create false edges.

4 Conclusion and Remarks

In this paper a novel automatic segmentation of left ventricle in T2 weighted MRI signals which is an indispensable step for quantifying oedema on T2 weighted MRI image combined geodesic morphological algorithm, fuzzy centre point detection, fuzzy multiresolution edge detection and convex hull to remove papillary muscle on endocardium is presented. The algorithm is then compared with manual segmentation by expert clinician and the initial results

are encouraging. The advantages of the proposed algorithm lies in its ability to overcome the effect of papillary muscles in the blood pool region along with its capability of tracing epicardium boundary even in the presence of weak edges information. Further work will be to classify oedema tissue within the LV wall and to create 3D visualisation for the LV and oedema areas.

Acknowledgement

The author would like to thank Golden Jubilee National Hospital, Glasgow for providing the CMR images.

References

1. Kellman, P., Aletras, A.H., Mancini, C., Mcveigh, E.R., Arai, A.E.: T2-prepared SSFP improves diagnostic confidence in oedema imaging in acute myocardial infarction compare to turbo spin echo. *Magn. Reson Med.* 57, 891–897 (2007)
2. Hsu, L.Y., Natanzon, A., Kellman, P., Hirsch, G.A., Aletras, A.H., Arai, A.E.: Quantitative myocardial infarction on delayed enhancement MRI Part I: Animal validation of an automated feature analysis and combined thresholding infarct sizing algorithm. *J. Magn. Reson. Imaging* 23, 298–308 (2006)
3. Baron, N., Kachenoura, N., Beygui, F., Cluzel, P., Grenier, P., Herment, A., Frouin, F.: Quantification of Myocardial Oedema and Necrosis during Acute myocardial Infarction. *Computer in Cardiology* (2008)
4. Stalidis, G., Maglaveras, N., Dimitriadis, A., Pappas, C., Strinzis, M.: Detection and Modeling of Infarcted Myocardium Region in MRI Image Using Contour Deformable Model. *Computer in Cardiology* (1995)
5. Ciofolo, C., Fradkin, M., Mory, B., Hauvast, G., Breeuwer, M.: Automatic Myocardium Segmentation in Late-Enhancement MRI. *ISBI* (2008)
6. Vincent, L.: Morphological grayscale reconstruction in image analysis: application and efficient algorithm. *IEEE Transaction on Image Processing* 2(2), 176–201 (1993)
7. Setahrehdan, S.K., Soraghan, J.J.: Automatic left ventricular centre point extraction in echocardiographic images. *Eurasip Signal Processing* 61(3), 275–288 (1997)

Research of Muscular Activity during Gait of Persons with Cerebral Palsy

Kristina Daunoravičienė¹, Jolanta Pauk²,
Jim Raso³, and Julius Griškevičius¹

¹ Vilnius Gediminas Technical University,
J. Basanavičiaus 28, LT-10225, Vilnius, Lithuania
e-mail: kristina.daunoraviciene@me.vgtu.lt,
bjulius.griskevicius@me.vgtu.lt

² Białystok Technical University,
Wiejska 45C, 15-351 Białystok
e-mail: jpauk@pb.edu.pl

³ Glenrose Rehabilitation Hospital,
Edmonton, Alberta, Canada
e-mail: jim.raso@capitalhealth.ca

Summary. Assessing muscular activity during gait in CP persons could provide valuable information in prescribing appropriate treatment to reduce the consequences of cerebral palsy as well as limiting further complication in cerebral palsy children. The main goal of this study was explored working regularities of muscle pairs in children population to show dependencies and variation on gait parameters. Functional evaluation was carried out on 20 cerebral palsy patients. The research have been done by using the system EMG. A surface electrode picked up on the main groups of muscles of lower limbs: the Rectus Femoris, the Vastus Lateralis, the Medial Hamstrings, the Lateral Gastrocnemius, and the Anterior Tibialis. There were several phases to the signal approach such as: data acquisition, data pre-processing, data modeling, data analysis and interpretation. From the results seen that for each task subjects have different strategies for keeping balance during walk depending on the basic level of muscle contraction or antagonistic and synergistic contraction required for that activity.

1 Introduction

EMG signal means an electrical signal acquired from any organ. It is usually a function of time, described in terms of amplitude, frequency and phase. The EMG signal measure electrical currents generated in muscles during its contraction representing neuromuscular activities [1, 2, 6]. Electromyography (EMG) is a one task to best representation of the neurological control of musculoskeletal system. During any voluntary movement such as gait, the

EMG signal is an algebraic summation of every muscle motor unit action rotations. So far, research and extensive efforts have been made in the area, developing better algorithms, upgrading existing methodologies, improving detection techniques to reduce noise, and to acquire accurate EMG signals. Observing activity of muscles during work could be evaluated by many parameters and methods. Results of such research are very useful for variety of fields, such as rehabilitation, ergonomics and etc., and have wide application [1, 2, 3, 4, 5, 6, 7, 9]. Most researchers assume normal gait symmetrical. Ounpuu and Winter assessed 7 muscles pairs from 10 subjects according to their dominant side. They determined that the assumption of symmetry is not valid for each subject [1, 2, 3, 4, 5]. The only muscles that had differences correlated to dominance, were the plantar flexors. Many scientific works note the presence of co-contraction due to overlapping of antagonistic activity during gait. However, these observations fail to quantify the extent of co-contraction. Such a measure is very difficult to quantify because the moment of force created by each antagonistic muscle would have to be known [4, 5, 6]. The contractions were quantified during walking, but this was limited only to the muscles of ankle joints [7]. The purpose of this work is to identify muscular activity parameters during walk of cerebral palsy subjects and show dependencies and variation of them on gait parameters.

2 Methodology

Functional evaluation was carried out on 20 cerebral palsy patients (ranged between 5 and 11 years), after clinical evaluation. The difficulties that the patients most commonly complained about were: climbing stairs, walking uphill, and bending down. Gait abnormalities of these persons were usually treated with a combination of rehabilitation, orthosis, and surgery. The subjects were analyzed while walking barefoot along a straight pathway. The research have been done by using the system EMG. A surface electrode picked up on the main groups of muscles of lower limbs: the Rectus Femoris, the Vastus Lateralis, the Medial Hamstrings, the Lateral Gastrocnemius, and the Anterior Tibialis with only minimal crosstalk from adjacent muscles. There were several phases to the signal approach such as: data acquisition, data pre-processing, data modeling, data analysis and interpretation. Based on the data from EMG, the dynamic model was created for detailed muscular activity during subjects gait analysis. It is not easy to analyze muscular activity with quite big distribution of values. That is why maximal amplitudes (A_{max}) of muscular activity indicated in EMG of left and right lower limbs in every stride were collected for all muscles under consideration of all subjects. Then maximal values from EMG were analyzed more features of normal muscular activity in pairs were noted. The difference between synergistic muscles' activity rates of this pair was evaluated us a ratio between maximal amplitudes of different muscles in every step:

$$R^S(A \max) = \frac{A \max_n^{S_1}}{A \max_{n+1}^{S_2}} \tag{1}$$

where n - number step, $A \max_n^{S_1}$, $A \max_{n+1}^{S_2}$ - maximal amplitude of appropriate muscle from analyzed pair. Whereupon the difference between synergistic muscles' activity rates of this pair was evaluated us a ratio between maximal amplitudes of different muscles in every step.

3 Results and Discussion

The dynamical model was created (Fig. 1) for detailed muscular activity during subjects gait analysis. Angles of kinematic links are not presented in the model and they don't represent magnitudes of movement torques and directions. This model presents upper body and consists of two main symmetrical parts representing right and left lower extremities. Lower limbs are expressed through elastic and stiff elements with their coefficients respectively for elasticity k and stiffness c for each segment of a limb and the same for right and left leg (Fig. 1a). Fig. 1b) represents one stage of gate - single support or the beginning of step phase (Fig. 2), which was analysed by evaluating muscular activity. Every elastic and stiff element in the model serves for a muscle pair analysed during walk. Firstly it was pair of synergists - *m. rectus femoris* and *m. medial hamstrings*, later antagonistic pair - *m. lateral gastrocnemius* and *m. anterior tibialis* was also evaluated.

Every stage of gate is very important during walk because muscles then works differently. For evaluating all of them information from two first steps of walking was taken and analyzed (Fig. 3,4).

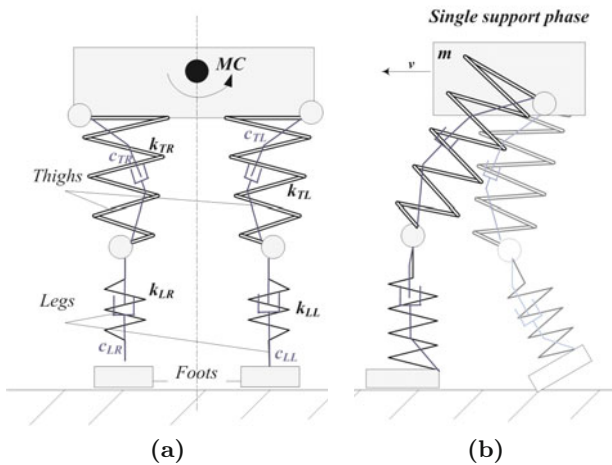


Fig. 1. Views of (a) analytical dynamic model, (b) model in the single support state

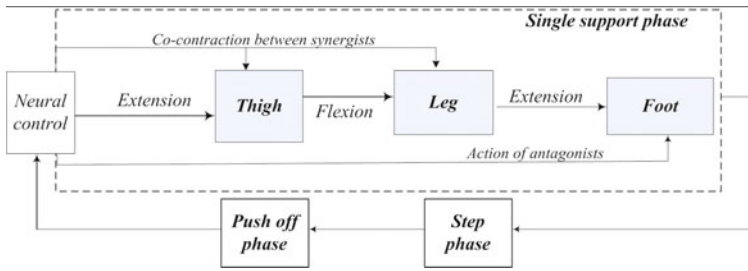
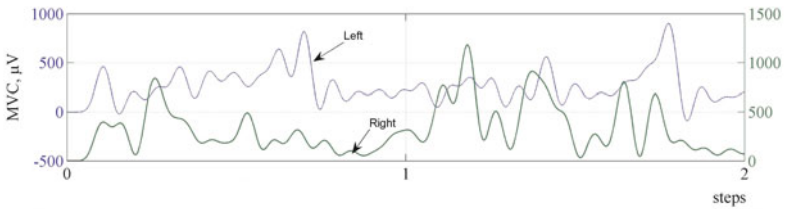
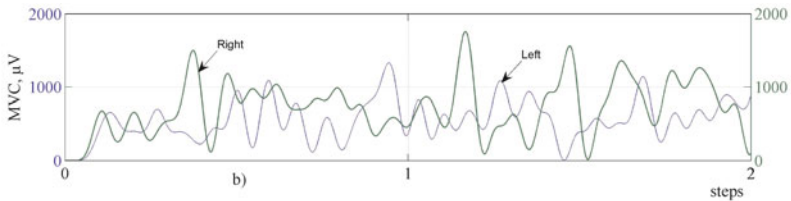


Fig. 2. Right/left lower limb movement scheme during single support phase



(a)



(b)

Fig. 3. Muscular activities of synergists during two steps of subject with CP walk: (a) activity of Medial Hamstrings, (b) activity of Rectus Femoris

The analytical cycle of subjects with spastic diplegia cerebral palsy is irregular and it could be noted from EMG data analysis in the muscular activity dependences as well as from simply visual gait analysis. From figure 3 it could be noticed that at the first step synergists show irregular activity without any co-contraction, therefore beginning of walk is unstable and with not strong support. At this part of gate biosystem was trying to adapt to circumstances and to evaluate all parameters for achieving movement. So in the second step some evidences of normal co-contraction could be found. Analyzing work regularities of antagonistic pair it was found strange thing. There were much more similarities in their activities either differences of them. It is not easy to analyze muscular activity from Fig. 3 and 4 with quite big distribution of values. That is why maximal amplitudes (A_{max}) of muscular activity indicated in EMG of left and right lower limbs in every stride were

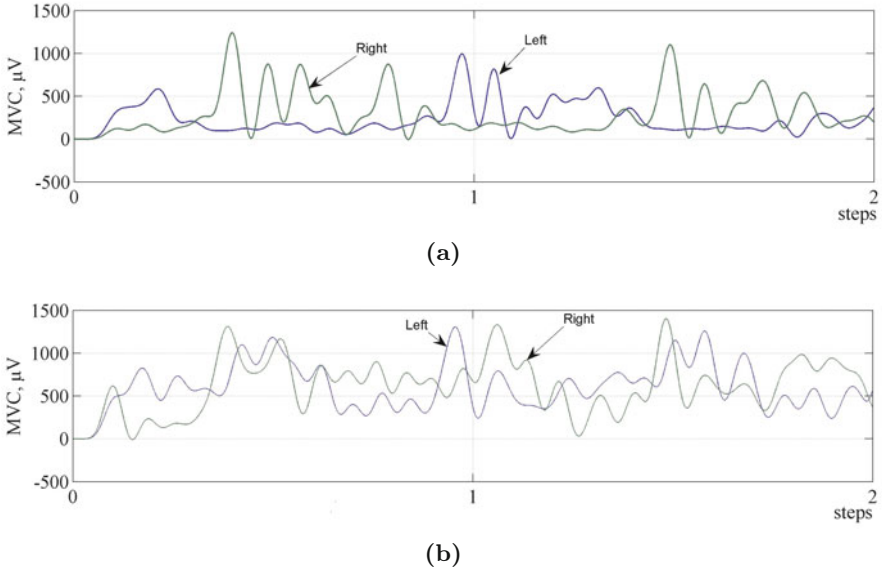


Fig. 4. Muscular activities of antagonists during two steps of subject with CP walk: (a) activity of Lateral Gastrocnemius, (b) activity of Anterior Tibialis

collected for all muscles under consideration of all subjects. And relations of these amplitudes on two steps of antagonistic and synergistic muscle pairs were obtained (Fig. 5). Then maximal values from EMG were analyzed more features of normal muscular activity in pairs were noted. For example in antagonistic pair (Fig. 5b) - when activity of one muscle is high, the value of another muscle activity is very low and in reverse order. But difference between decrease and increase of muscular activity vary a lot and got in the range $140 < A < 1008\mu V$. Variations of activity differ from each other in right and left leg, for left leg these values vary approximately in the interval $200 < A < 750$. From Fig. 5a the same delay which was mentioned before in co-contraction was found during first step. Differences between muscular activity delays were evaluated in about a half of trail and were also expressed in time duration percents not only in maximal values. Approximately 20 % of time delay in all biosystems was noted, sometimes it exceeds till 50 %. Such a situation turns in big energy dissipation. This value varies according to the system state and wellness. Maximal amplitudes of Lateral Gastrocnemius, and the Anterior Tibialis EMG signals variation in every step presented in Fig. 5a shows work regularities of antagonists' pair as well as in the 5a shows activities of right and left leg synergistic muscles' pair. Likewise it is not so easy to define the difference which must be between them normally, because it is individual rate. Ratio calculated according expression (1) presented in the Fig. 6. Normally ratio in synergists' pair should show values approaching to 1, if they work properly during walk. But these values distribute in the

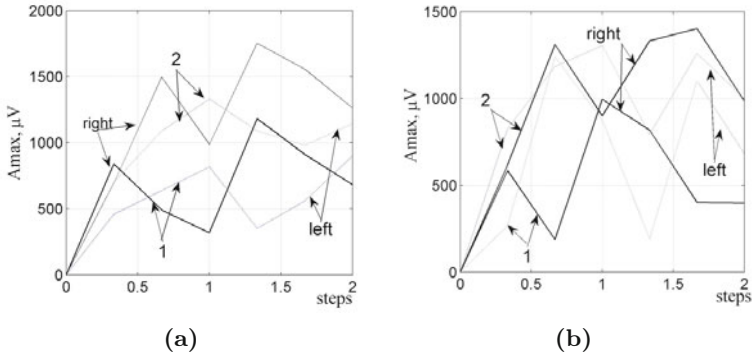


Fig. 5. Maximal amplitudes of muscles pair activity of one subject with cerebral palsy: (a) synergistic muscles pair where 1 - maximal amplitudes of Medial Hamstrings, 2 - maximal amplitudes of Rectus Femoris, (b) antagonistic pair where 1 - maximal amplitudes of Lateral Gastrocnemius, 2- maximal amplitudes of Anterior Tibialis

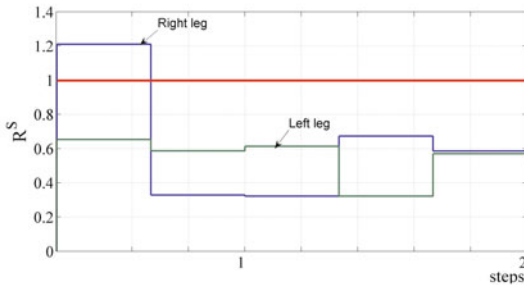


Fig. 6. Ratio between A_{max}^S of Medial Hamstrings and Rectus Femoris synergists muscles

interval $0.5 \leq R_A \leq 1.5$ The ratio R could be calculated not only for evaluation of synergists' muscles activity during movements but also for assessment of inequality between one subject left and right limb muscles' activity.

4 Conclusions

From the results seen that for each task subjects have different strategies for keeping balance during walk depending on the basic level of muscle contraction or antagonistic and synergistic contraction required for that activity. Because synergist and antagonists perform different work and energy they could be evaluated by different methods. Possible evaluation of them is ascertaining of ratios. For synergists this ratio varies approximately $0.35 \leq R^S \leq 1.2$, so it deviates from normal ratio about two times during gait of subject with spastic diplegia cerebral palsy. Antagonists activity showed irregularities and

the rates differ from each other more than two times. Delay in biosystem was noted during co-contraction of synergy muscles and it varies from 20 to 50 % until their proper working. Such biosystem has big energy dissipation. Noticed, that muscle pairs analyzed in the paper can't produce normal activity and cover each other during normal gait cycle, and therefore movements are overbalanced. Results of analyzed subsystems showed that pairs of different acting muscles influence not only gait parameters and stability balance but also the main efficiency of biosystem.

References

1. Hobará, H., Kanosue, K., Suzuki, S.: Changes in muscle activity with increase in leg stiffness during hopping. *Neuroscience Letters* 418, 55–59 (2007)
2. Hortobágyi, T., Solnik, S., Gruber, A., Rider, P., Steinweg, K., Helseth, J., Devita, P.: Interaction between age and gait velocity in the amplitude and timing of antagonist muscle coactivation. *Gait & Posture* 29(4), 558–564 (2009)
3. Yamazaki, Y., Suzuki, M., Ohkuwa, T., Itoh, H.: Coactivation in arm and shoulder muscles during voluntary fixation of a single joint. *Brain Research* 59(6), 439–446 (2003)
4. Schmitz, A., Silder, A., Heiderscheit, B., Mahoney, J., Thelen, D.G.: Differences in lower-extremity muscular activation during walking between healthy older and young adults. *Journal of Electromyography and Kinesiology*, 1050–6411 (2008)
5. Shimada, H., Suzuki, T., Kimura, K., Hirata, T., Sugiura, M., Endo, Y., Yasuhara, K., Shimada, K., Kikuchi, K., Oda, K., Ishii, K., Ishiwata, K.: Effects of an automated stride assistance system on walking parameters and muscular glucose metabolism in elderly adults. *British Journal of Sports Medicine* 42, 922–929 (2008)
6. Ounpuu, S., Winter, D.A.: Bilateral electromyographical analysis of lower limbs during walking in normal adults. *EEG Clin. Neurophysiology* 72, 429–438 (1989)
7. Falconer, K., Winter, D.A.: Quantitative assessment of co-contraction at the ankle joint in walking. *Electromyogr. & Clin. Neurophysiol.* 25, 135–149 (1985)
8. Pauk, J.: Different techniques for EMG signal processing. *Journal of Vibroengineering* 10(4), 571–576 (2008)
9. Lloyd, A.: EMG Driven Musculoskeletal Modelling: Finding Out How Muscles Control Movement. In: IX International Symposium on Computer Simulation in Biomechanics, pp. 9–10 (2003)

Automated Recognition of Abnormal Structures in WCE Images Based on Texture Most Discriminative Descriptors

Piotr Szczypiński and Artur Klepaczko

Technical University of Łódź, Institute of Electronics,
ul. Wolczanska 211/215, 90-924 Łódź, Łódź
e-mail: {pms, aklepaczko}@p.lodz.pl

Summary. In this paper we study the problem of classification of wireless capsule endoscopy images (WCE). We aim at developing a computer system that would aid in medical diagnosis procedure. The goal is to automatically detect images showing pathological alterations in an 8-hour-long WCE video. We focus on three classes of pathologies, ulcers, bleedings and petechia, since they are typical for several diseases of intestines. Utilized are methods for image texture and color analysis to obtain numerical description of images. Then, three methods for selection of most discriminative descriptors are used, namely Vector Supported Convex Hull, Support Vector Machines and Radial Basis Function Networks. The results produced by the three methods are compared.

1 Introduction

Wireless capsule endoscopy (WCE) is a non-invasive technique which has ability for visualization of the whole human small intestine. However, interpretation of a video sequence produced by the capsule endoscope demands significant effort and remains a time consuming procedure. This research aims at development of a system that would allow automatic detection of WCE video frames showing pathological changes.

The adopted approach utilizes texture and color analysis of images to numerically describe thoroughly selected image fragments. . In the literature ([5](#), [17](#)), one can find several approaches to texture description. It is assumed that texture and color descriptors (features) enable differentiation between normal and pathologically altered tissues. The main problem which arises here is the multidimensional feature space generated by the analysis. It is difficult to arbitrarily choose a particular model representative for pathological tissue color and texture. Therefore, in the presented approach we compute a large number of various descriptors. Then, the subset of several most discriminative descriptors is selected.

The methods for descriptors selection aim at finding a subspace of descriptors space, which at best satisfies some chosen classification rule. The aim of classification is to separate images with pathologies present from other images. We assume the subspace exists, that descriptor vectors of the pathology class form a cluster surrounded by other vectors. Therefore, we use methods capable to find such subspaces. There are three methods used for the subset selection purpose, Support Vector Machines (SVM) with the radial kernel, Radial Basis Function Networks (RBFN) and the formulated by the authors the method of Vector Supported Convex Hull (VSCH) [14].

The remainder of this paper is structured as follows. Section 2 reviews the main aspects of the WCE imaging technique. The notion of texture analysis and the software tools used in this research are described in Section 3. In Section 4 we recall general principles of the feature subspace selection methods. In the experimental part (Section 5), efficiency of VSCH is tested in comparison with the SVM and RBFN. Finally, Section 6 concludes.

2 Capsule Endoscopy

Wireless capsule endoscopy (WCE) [6, 12], is a technique that facilitates the imaging of the human gastrointestinal system including small intestine. The WCE system consists of a pill-shaped capsule (cf. Fig. 1a) with built-in video camera, light-emitting diodes, video signal transmitter and battery, as well as a video signal receiver-recorder device. The wireless capsule endoscope used in this study produces color images of the internal lumen (cf. Fig. 1b). The images cover a circular 140° field of view. A patient under investigation ingests the capsule, which then passes through the gastrointestinal tract. When the capsule goes through the small bowel it is propelled by peristaltic movements. The capsule transmits video data at a rate of two frames per second for approximately 8 hours. Investigation of the recorded video is performed by a trained clinician. It is a tedious task that usually takes more than an hour. The video interpretation involves viewing the video and searching for abnormal-looking entities like bleedings, erosions, ulcers, polyps and narrowed sections of the bowel.

There is a need for automatic methods which would aid in the investigation process, either by reduction of the time spent on the process or by focusing the attention of the clinician on medically relevant video fragments. Several research groups have recently reported methods for gastro-intestinal tract segmentation into sections utilizing various color and texture descriptors (e.g. in [3, 4, 8, 9, 10]). Another classification procedure is presented in [18]. Image features, such as intensity, contrast, and selected geometrical parameters, are extracted from WCE video frames. By comparison of feature vectors computed for nearby frames, video fragments presenting contractions can be detected. Another approach is to obtain an image of the bowel surface [15] by preprocessing the video. Such an image, a bowel map, enables to quickly

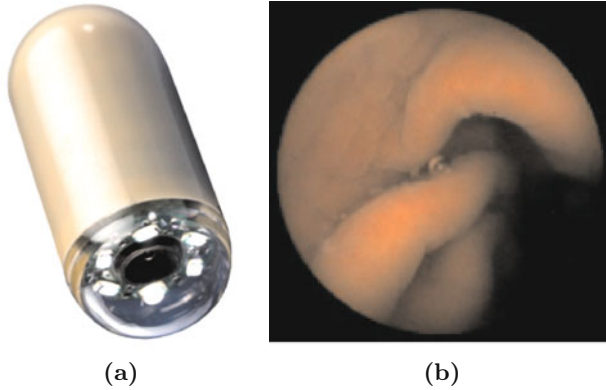


Fig. 1. Wireless capsule endoscope (a) and an example image (b)

get an overview of the entire recording in terms of completeness and quality and presence of certain abnormalities.

In the presented approach we presume that image regions containing different pathologies and various aspects of normal mucosal appearance also differ in terms of color and texture descriptors. Such descriptors can be computed and then used for differentiation of image contents.

3 Texture Analysis

A texture can be viewed as a complex composition of spatially organized, repeated patterns, which are perceived to demonstrate specific brightness, color, size, roughness, directivity, randomness, smoothness, granulation, etc. A texture may carry substantial information about the structure of physical objects — human tissues or organs in particular. Consequently, textural image analysis is an important issue in image processing and understanding in medical applications. To perform such analysis, mathematically defined texture properties are computed.

In our study we use Szczypinski:13 4.7 software [13, 16] for textural feature computation. The software is capable of conducting a quantitative analysis of texture within arbitrarily selected regions of interest (ROI). There are three categories of feature computation approaches that Szczypinski:13 utilizes: statistical (based on image histogram, gradient, co-occurrence matrix, run-length matrix), model-based (implementation of the autoregressive model) and image transform (based on the Haar wavelet). Szczypinski:13 may be used to compute textural descriptors based on color components of a color spaces, such as RGB, YUV, YIQ, HSB, XYZ, L^*a^*b . The textural features computed for different color components can be combined to obtain a comprehensive characterization of a colored texture. Therefore, feature vectors computed by Szczypinski:13 may include even 3 thousands of elements per

individual region of interest. The need for dimensionality reduction becomes indispensable.

4 Vector Supported Convex Hull Method

The VSCH is a discriminant analysis method of supervised learning for reduction of vectors dimensionality and for data classification. Usually only a limited number of features carry relevant information needed for discrimination and other descriptors are redundant for classification. The VSCH aims at finding a subspace of descriptors (subset of descriptors), which present best discrimination ability to separate two classes (sets) of vectors. Moreover the VSCH produces a classification rule to separate the two classes.

To explain the concept of VSCH let us assume input data consist of two sets (classes) of feature vectors in an n -dimensional space. All the features are real numbers. We search for a k -dimensional subspace ($k < n$) such that vectors of the set number one form a cluster surrounded by vectors of the set number two. Let us consider a convex hull of set one in a k -dimensional subspace of feature vectors space. There are several algorithms for finding convex hull of sets of vectors. In our implementation we use a quick-hull algorithm [1].

Now we define a coefficient Q_1 . It is the number of vectors belonging to the second class, which also belong to the convex hull built on class number one. The next step is to find a centroid (or gravity center) of the convex hull. Then the convex hull is isotropically scaled up around the fixed centroid. We find the maximum scaling factor a for which Q_1 does not increase. Now we define a Q_2 coefficient which is equal to reciprocal of a . Since, a is larger than 1, the coefficient Q_2 is a fraction. On the other hand coefficient Q_1 is an integer number equal or higher than 0. Now, we combine the two coefficients and define a comprehensive Q coefficient as:

$$Q = Q_1 + Q_2. \quad (1)$$

The Q specifies discriminative power of k -dimensional feature space. The lower value of the Q indicates the analyzed feature subspace has better class separability.

The algorithm for feature space reduction based on VSCH searches through all the 1D, 2D, 3D and 4D feature subsets and computes Q coefficient for each subset. For further analysis and classification purpose such subset is chosen, which exposes the lowest Q coefficient. The algorithm also produces rules for classification. The rules are given by a set of inequalities defining the convex hull scaled up by factor of \sqrt{a} .

In many medical applications it is crucial not to overlook any indications of pathology. Such indications usually are later verified by medical experts and may be rejected. If they are mistakenly rejected by an automatic method, an expert may never notice the pathology. Therefore, it is important to find

methods characterized by a minimal false negative error. The VSCH method reveals a property which is particularly useful in biomedical image analysis. The method produces classification rules, for which (for the training set vectors) the false negative error is equal to zero. The minimization of false positive errors is a secondary goal, and is achieved directly by minimization of the Q_1 coefficient.

5 Experiment

In the experimental part of this study we analyzed a set 20 WCE video sequences from which we selected 30 frames showing different forms of bleedings, 50 frames with ulcerations, 100 images of petechia and over 600 images with no visible pathology structures. The latter group served as a reference class in the analyzed discrimination problems. Regions of bleeding, ulceration and petechia (regions of interest) were manually depicted within the images. Then all the selected frames were divided into circular overlapping subregions, each of 2009 pixels area (cf. Fig. 2).

For images showing abnormal changes, textural features were computed within circular subregions enclosed within the depicted regions of interest. For reference images textural features were computed within circular subregions enclosed within the whole image field of view. The number of calculated feature vectors amounted to 128 for each pathology and 1024 in the case of reference images. These vectors were then combined into three 2-class data sets, each containing one class representing abnormal tissues and the reference group. The data sets were first submitted to the VSCH method and also to feature selection and classification procedures that employed SVM and RBF network algorithms. In the two latter cases, feature space exploration was performed using exhaustive search [2, 7, 11] and relevance of a feature subset

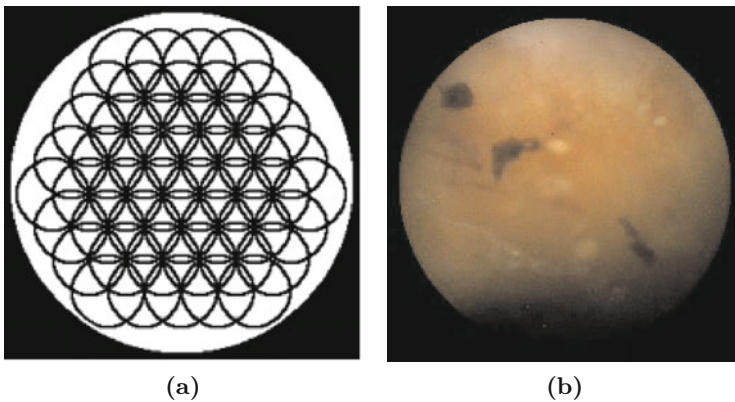


Fig. 2. Overlapping circular regions used in the study (a) and example image of petechia class

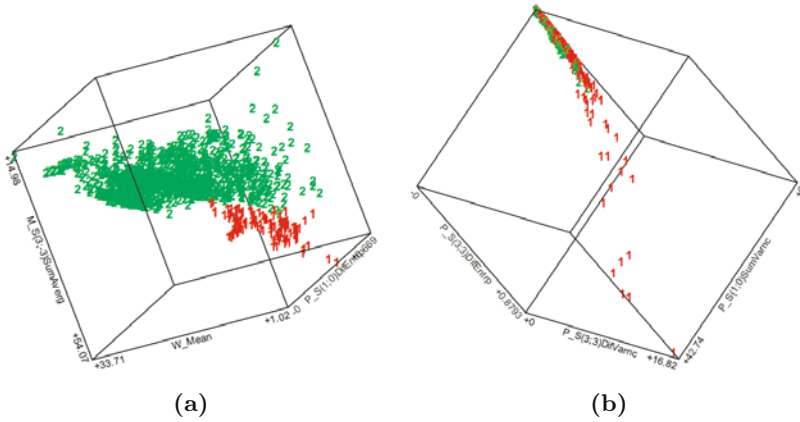


Fig. 3. Scatter plots of data vectors in the best feature spaces selected by the VSCH method: bleeding vs. normal (a) and petechia vs. normal (b)

Table 1. Classification results

		FPR [%]	Specificity	FNR [%]	Sensitivity
VSCH	bleeding vs. normal	9.3	0.907	0.0	1.000
	petechia vs. normal	3.4	0.966	0.0	1.000
	ulcer vs. normal	2.5	0.975	0.0	1.000
SVM	bleeding vs. normal	2.3	0.977	0.8	0.999
	petechia vs. normal	6.4	0.936	1.6	0.984
	ulcer vs. normal	4.3	0.958	6.4	0.936
RBFN	bleeding vs. normal	6.3	0.937	0.3	0.999
	petechia vs. normal	10.9	0.891	1.9	0.981
	ulcer vs. normal	3.9	0.961	9.0	0.910

was assessed by the accuracy of a corresponding classifier. The classification specificity, sensitivity, false positive rate and false negative rate computed for the training data sets are presented in Table 1. Example scatter plots of feature vectors in the VSCH-selected texture parameter spaces are viewed in Fig. 3.

6 Results Discussion and Conclusions

Analysis of the obtained results leads to the following conclusions. First of all, the performed experiments confirm that texture analysis provides a practical numerical description of the WCE images. It is possible to classify different types of visualized tissues basing on the selected, most relevant texture parameters. Among the calculated attributes, color component features appear

to be the best at discriminating abnormal and normal regions. Secondly, the error rates as well as accuracy measures viewed in Table 1 are comparable for all the tested approaches to feature selection.

The VSCH method appears to be overoptimistic when predicting the False Negative Ratio (FNR) on the training set. This results directly from the very nature of the algorithm which aims at construction of a convex hull around all vectors from a chosen pathology class. We expect this error to be higher than zero if other set of vectors is used for verification.

Usage of both SVM and RBFN involves problem-specific parameterization of a kernel function. Frequently, one must experiment with several values of power exponents (both in polynomial or radial basis functions) before a final choice can be made. On the other hand, VSCH is a non-parametric method and does not require any fine-tuning to solve particular tasks. Moreover, it does not require any feature space standardization. Also any other linear transformation of feature space has no influence on the result produced by the method.

Acknowledgments

This work was supported by the Polish Ministry of Science and Higher Education grant no. 3263/B/T02/2008/35.

References

1. Barber, C., Dobkin, D., Huhdanpaa, H.: The quickhull algorithm for convex hulls. *The Quickhull algorithm for convex hulls* 22(4), 469–483
2. Blum, A.L., Langley, P.: Selection of relevant features and examples in machine learning. *Artificial Intelligence* 97, 245–271 (1997)
3. Coimbra, M., Campos, P., Cunha, J.: Extracting clinical information from endoscopic capsule exams using mpeg-7 visual descriptors. In: *Integration of Knowledge, Semantics and Digital Media Technology, EWIMT 2005*, pp. 105–110 (2005)
4. Coimbra, M., Cunha, J.: Mpeg-7 visual descriptors-contributions for automated feature extraction in capsule endoscopy. *IEEE Transactions on Circuits and Systems for Video Technolog* 16(5), 628–637 (2006)
5. Haralick, R.: Statistical and structural approaches to texture. *IEEE Proceedings* 67(5), 768–804 (1979)
6. Iddan, G., Meron, G., Glukhowsky, A., Swain, P.: Wireless capsule endoscopy. *Nature* 405(6785), 417–418 (2000)
7. Kohavi, R., John, G.H.: Wrappers for feature subset selection. *Artificial Intelligence* 97, 273–324 (1997)
8. Li, B., Meng, M.H.: Ulcer recognition in capsule endoscopy images by texture features. In: *7th World Congress on Intelligent Control and Automation, WCICA 2008*, pp. 234–239 (2008)

9. Mackiewicz, M., Berens, J., Fisher, M.: Wireless capsule endoscopy video segmentation using support vector classifiers and hidden markov. In: Proceedings of the International Conference on Medical Image Understanding and Analyses (2006)
10. Mackiewicz, M., Berens, J., Fisher, M., Bell, G.: Colour and texture based gastrointestinal tissue discrimination. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, vol. 2, pp. 597–600 (2006)
11. Pudil, P., Somol, P.: Current feature selection techniques in statistical pattern recognition. In: Kurzynski, M., Puchala, E., Wozniak, M., Zolnierok, A. (eds.) Computer Recognition Systems, Advances in Sof. Computing, vol. 30, Springer, Heidelberg (2005)
12. Swain, P., Fritscher-Ravens, A.: Role of video endoscopy in managing small bowel disease. *GUT* 53, 1866–1875 (2004)
13. Szczypinski, P.: (2009), <http://www.eletel.p.lodz.pl/MaZda> (visited: May 2010)
14. Szczypinski, P., Klepaczko, A.: Convex hull-based feature selection in application to classification of wireless capsule endoscopic images. In: Blanc-Talon, J., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS 2009. LNCS, vol. 5807, pp. 664–675. Springer, Heidelberg (2009)
15. Szczypinski, P., Sriram, R.D., Sriram, P., Reddy, D.: A model of deformable rings for interpretation of wireless capsule endoscopic videos. *Medical Image Analysis* 13(2), 312–324 (2009)
16. Szczypinski, P., Strzelecki, M., Materka, A., Klepaczko, A.: Mazda - a software package for image texture analysis. *Computer Methods and Programs in Biomedicine* 94, 66–76 (2009)
17. Tuceryan, M., Jain, A.: Texture Analysis. In: *The Handbook of Pattern Recognition and Computer Vision*, pp. 207–248. World Scientific Publishing Co., Singapore (1998)
18. Vilarinao, F., Kuncheva, L.I., Radeva, P.: Roc curves and video analysis optimization in intestinal capsule endoscopy. *Pattern Recogn. Lett.* 27(8), 875–881 (2006)

Augmented Reality Interface for Visualization of Volumetric Medical Data

Tomasz Hachaj¹ and Marek R. Ogiela²

¹ Pedagogical University of Cracow,
Institute of Computer Science and Computer Methods,
2 Podchorążych Ave, 30-084 Kraków, Poland
e-mail: tomekhachaj@o2.pl

² AGH University of Science and Technology, Institute of Automatics
30 Mickiewicza Ave, 30-059 Kraków
e-mail: mogiela@agh.edu.pl

Summary. Most of the publications in the field of augmented reality in medicine concentrates on developing new types of systems but not on optimization of speed or portability of existing ones. The intention of authors of this article is to fill this gap by showing a possibility of creating augmented reality interface for visualization of volumetric medical data for Windows OS environment that can be run on off - the - shelf computer and test its capability. Tests of the authors application was performed on three 3D models generated from computer tomography data stored in collection of DICOM files. Authors novel and efficient solution can be easily attached to nearly all medical application system running Windows OS and give medical personnel support of a new level by presenting more informative and realistic three dimensional visualizations for low cost.

1 Introduction

Augmented reality (AR) is a technology that allow the real time fusion of computer generated digital content with the real world. Unlike virtual reality (VR), that completely immerses users inside a synthetic environment, augmented reality allows the user to see three dimensional virtual objects superimposed upon the real word [1].

Analyzing the articles and conference proceedings of the leading AR research symposium several significant research directions on this field can be identified; those are: tracking techniques, display technologies, mobile augmented reality and interaction techniques.

Tracking techniques is a group of computer methods that are used to achieve robust and accurate overlay of virtual imagery. There are various optical tracking techniques including fitting a projection of a 3D model onto detected features in the video image, matching a video frame with photos and real time image registration algorithms [2]. One of the most popular

method is a calculation of the 3D pose of a camera relative to a single square tracking marker. The commonly used display technologies are head mounted, handheld, and projection displays for AR. Many contemporary mobile devices like iPods and cell phones can execute sophisticated algorithms and are equipped with high resolution color displays and cameras. For those devices AR applications for outdoor usage are developed. The last type of methods is interaction techniques that enables user interaction with AR content.

Augmented reality shows its usefulness especially in the field of the medicine [3]. The most notable examples are deformable body atlases, AR surgical navigation systems, interfaces and visualization systems. Augmented Reality aim at lifting the support to pre, intra and post - operative visualization of clinical data a new level by presenting more informative and realistic three dimensional visualizations [4]. Not many years ago medical AR systems required expensive hardware configuration and dedicated OS to handled real time superimposing and visualization process [2]. Most of the publications in the field of AR in medicine concentrates on developing new types of systems but not on optimization of speed or portability of existing ones. The intention of authors of this article is to fill this gap. The authors goal was to create portable augmented reality interface for visualization of volumetric medical data for Windows OS environment that can be run on off - the - shelf computer and test its capability. The main advantage of proposed solution is speed, quality of rendering of virtual object and low hardware cost.

In order to achieve that goal the three basic conditions must be satisfied. At first software must be written in managed code running on the virtual machine in order to be independent of hardware installed and version of Windows OS. The second condition is capability of detection position of virtual object in the real scene in the real time. The third condition is the real time rendering of volumetric medical data. All of this conditions was satisfied by authors solution based on .NET platform, NyARTToolkit CS and XNA Framework 2.0 (the last two components will be described in the following paragraphs). Authors implemented the solution with C# language.

In order to run application end user needs NET Framework 2.0 SP1 (this framework is included in new distributions of windows OS), DirectX 9.0c library and XNA Framework Redistributable 2.0. All of this components can be downloaded free of charges.

2 Augmented Reality Environment

For identification of the position of virtual object in the real scene authors used size - known square markers. The transformation matrices from these marker coordinates to the camera coordinates are estimated by image analysis. After thresholding of the input image, regions whose outline contour can be fitted by four line segments are extracted. Parameters of these four line segments and coordinates of the four vertices of the regions found from the intersections of the line segments are stored for later processes. The regions

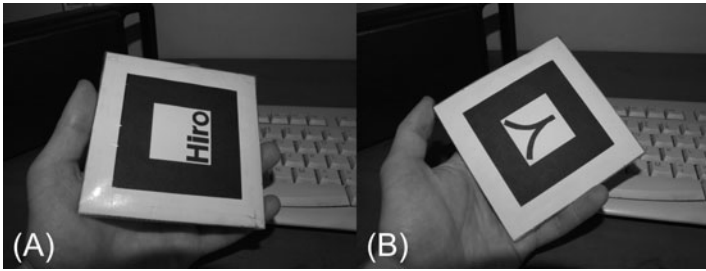


Fig. 1. Two image patterns (ARToolkit markers) used for AR software printed on both sides of cardboard. Over the second marker virtual objects are rendered upside - down.

are normalized and the sub-image within the region is compared by template matching with patterns that were given the system before to identify specific user ID markers. The further details of the algorithm can be found in [1]. The marker for the authors system is a cardboard with two image patterns printed on both sides (Fig. 1). Over the second pattern virtual objects are rendered upside - down in order to enlarge the possible field of observation of the virtual objects.

Authors used NyARTToolkit CS [5] that is the implementation of well known ARToolkit [6] in C# programming language. The important fact is that this module uses no native code (only managed).

The NyARTToolkit CS is dedicated for DirectX that uses left handed coordinate system, so the position of marker is defined by transformation matrix for that type of system. Because XNA uses the Right-handed coordinate system it is necessary to compute transformation coefficients from one system to another. In order to compute translation and rotation in 3D one uses 3 coordinates for translation and 3 coordinates for rotation. Changing transform matrix from one system to another can be simply done by changing the sign of translation coefficients and rotation angles:

$$[tx, ty, tz, rx, ry, rz]_L = [tx, ty, -tz, -rx, -ry, rz]_R \quad (1)$$

Where tx, ty, tz are translation coefficients and rx, ry, rz are angles of rotation around x, y and z axis.

3 Real Time Rendering of 3D Data

There are two most popular methods capable of real time rendering 3D volume data: texture based volume rendering and volume ray casting. Authors choose the ray casting GPU accelerated algorithm described in [7]. Algorithm uses standard front to back blending equations in order to find color and opacity of rendered pixels:

$$C_{dst}(i+1) = C_{dst}(i) + (1 - \alpha_{dst}(i)) \alpha_{src} C_{src} \quad (2)$$

$$\alpha_{dst}(i+1) = \alpha_{dst}(i) + (1 - \alpha_{dst}(i)) \alpha_{src} \quad (3)$$

Where C_{dst} and α_{dst} are the color and opacity values of the rendered pixels and C_{src} and α_{src} are the color and opacity values of the incoming fragment.

The approach proposed in [7] includes standard acceleration techniques for volume ray casting like early ray termination and empty-space skipping. By means of these acceleration techniques, the framework is capable of efficiently rendering large volumetric data sets including opaque structures with occlusions effects and empty regions. For many real world data sets, the proposed method is significantly faster than previous texture based approaches, yet achieving the same image quality.

The algorithm was implemented in XNA Framework with High Level Shader Language (HLSL). XNA Framework was originally designed as a set of tools that facilitates computer game development and management for Windows an Xbox. The basic window frame of XNA Framework is called "Game" and authors used this notation in the next paragraph.

4 The System Performance Test

Authors solution was tested on off - the - shelf computer with Intel Core 2 Duo CPU 3.00 GHz processor, 3.25 GB RAM, Nvidia GeForce 9600 GT graphic card with 32 - bit Windows XP Professional. Creative pd 1120 USB web cam was used as a video capture device.

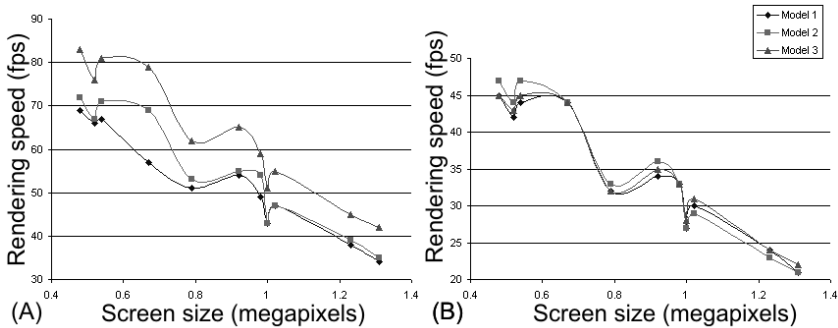
The speed of marker detection algorithm was 30 frames per second (fps) for 320×240 resolution of the camera and 15 fps for 640×480 . Those are the maximal speed of image capture for authors web camera so the fps limit of marker detection software could not be found without better (faster) camera.

The speed of 3D rendering depends of the size of rendered object. The smaller the object is (less number of pixel to compute) the rendering process runs faster. In authors research we localize the virtual object in the nearest position to the camera as it is possible in order to generate the maximal possible size of the rendered model. Tests was performed on three 3D models generated from computer tomography (CT) data acquired from three different patients. Data was stored in collection of DICOM files. The size of rendered volume was $256 \times 256 \times 248$ pixels, $256 \times 256 \times 240$ pixels and $256 \times 256 \times 207$ pixels. Each model was rendered in two modes: with no semi transparent pixels and with semi transparent pixels. Semi transparent pixels was used for simulation tissues with low density (in example skin). The speed of rendering was measured (in frames per seconds) for different popular screen resolution. The results are presented in Table 1 and in Fig. 2.

As can be seen rendering speed decreases with size of rendered object but not necessarily with resolution of the screen. In example for model 1 with no semi transparent pixels for 1280×768 we have 49 fps, for 1152×864 43 fps

Table 1. Dependents of speed of rendering three different 3D models on XNA Game window size (megapixels)

XNA Game resolution (megapixels)	Rendering speed (fps)					
	No semi transparent pixels			With semi transparent pixels		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
800 × 600 (0.48)	69	72	83	45	47	45
832 × 624 (0.52)	66	67	76	42	44	43
960 × 600 (0.54)	67	71	81	44	47	45
1088 × 612 (0.67)	57	69	79	44	44	44
1024 × 768 (0.79)	51	53	62	32	33	32
1280 × 720 (0.92)	54	55	65	34	36	35
1280 × 768 (0.98)	49	54	59	33	33	33
1152 × 864 (1.00)	43	43	51	27	27	28
1280 × 800 (1.02)	47	47	55	30	29	31
1280 × 960 (1.23)	38	39	45	24	23	24
1280 × 1024 (1.31)	34	35	42	21	21	22

**Fig. 2.** Data from Table 1 presented in the charts. Dependents of speed of rendering three different 3D models (fps) on XNA Game window size (megapixels) (A) models with no semi transparent pixels, (B) models with semi transparent pixels

and for 1280×800 47 fps. This is because model in 1152×864 is bigger (visualized with higher number of pixels) then in two other cases.

For authors hardware the best configuration of application window size / camera resolution and capture speed is 1024×768 (or 1280×768) for window and 320×240 (30 fps) for camera captured image size or 1280×1024 and 640×480 (15 fps). In those two configurations all detected markers position will be instantly computed to proper linear transformation of virtual 3D model and rendered.

The speed of rendering was higher for models with no semi transparent pictures because they required less computation time. The example visualization performed by authors solution is presented in Fig. 3.



Fig. 3. Example augmented reality visualization of three different models constructed from CT data. In the first row model with semi transparent pixels for tissues with low density. In the second row model with no semi transparent pixels. In the third row visualization of dense tissues (bones).

5 Conclusion

Authors created the novel and efficient library that delivers augmented reality interface for visualization of volumetric medical data for .NET platform. The software enables interface to run in real time (about 30 fps) on off - the - shelf computer with resolution 1024×768 . The limitation of marker detection algorithm speed is an effect of reaching the maximal speed of image capture for authors USB web camera. This might be overcome by purchase professional camera with fast data transfer connector to the PC.

Authors solution can be easily attached to nearly all medical application system running on Windows OS and give medical personnel support of a new level by presenting more informative and realistic three dimensional visualizations for low cost.

There are also some limitation of the software that authors plan to eliminate in the further versions of the solution. Because two dimensional markers was utilized there are some angles in which AR marker becomes invisible or impossible to indentify. This problem might be eliminated by using three dimensional (in example cube) markers with standard ARToolkit patterns stick to each side of them, or different type of three dimensional markers not based on ARToolkit [8].

The second limitation is a speed of the rendering algorithm used by authors. For the high resolution of screen it might be necessary to apply another (faster) versions of rendering algorithm.

The last problem might be running the software on non Windows OS. There is a possibility that the problem might be solved by using cross platform, open source .NET and XNA framework (like Mono with Mono.Xna) but the compatibility of those requires further researches.

Authors are planning to use their library for visualization purpose in the computer - assisted diagnosis systems of their design [9, 10].

References

1. Haller, M., Billinghurst, M., Thomas, B.: *Emerging Technologies of Augmented Reality: Interfaces and Design*. Idea Group Publishing, USA (2006)
2. Warfield, S., et al.: *Advanced Nonrigid Registration Algorithms for Image Fusion*. *Brain Mapping: The Methods*, 2nd edn., pp. 661–690. Academic Press, San Diego (2002)
3. Yang, G., Jiang, T.: *Medical Imaging and Augmented Reality*. Second International Workshop. In: *MIAR 2004* (2004)
4. Denis, K., et al.: *Integrated Medical Workflow for Augmented Reality Applications*. In: *International Workshop on Augmented environments for Medical Imaging and Computer-aided Surgery, AMI-ARCS* (2006)
5. NyARTToolkit CS home page, <http://nyatla.jp/nyartoolkit/wiki/index.php>
6. Kato, H., Billinghurst, M.: *Marker Tracking and HMD Calibration for a video-based Augmented Reality Conferencing System*. In: *Proceedings of the 2nd International Workshop on Augmented Reality (IWAR 1999)*, pp. 85–94 (1999)
7. Krüger, J., Westermann, R.: *Acceleration Techniques for GPU-based Volume Rendering*. *IEEE Visualization* (2003)
8. Marmulla, R., et al.: *An augmented reality system for image-guided surgery*. *Medical Image Analysis* 34 (2005)
9. Ogiela, L., Tadeusiewicz, R., Ogiela, M.R.: *Cognitive Methods in Medical Image Analysis and Interpretation*. In: *4th International Workshop on Medical Imaging and Augmented Reality, MIAR 2008, Proceedings on CD ROM*, paper P1-12, pp. 1–7. he University of Tokyo, Tokyo (2008)
10. Hachaj, T., Ogiela, M.R.: *Computer - assisted diagnosis system for detection, measure and description of dynamic computer tomography perfusion maps*. In: Choraś, R.S., Zabłudowski, A. (eds.) *Image Processing and Communication Challenges*, pp. 350–357 (2009)

Biomedical Computer Vision Using Computer Algebra: Analysis of a Case of Rhinocerebral Mucormycosis in a Diabetic Boy

Mario Vélez and Juan Ospina

Logic and Computation Group, Physical Engineering Program,
School of Sciences and Humanities,
EAFIT University Medellín, Colombia
e-mail: {mvelez, jospina}@eafit.edu.co

Summary. Computer algebra is applied to biomedical computer vision. Specifically certain biomedical images resulting from a case of rhinocerebral mucormycosis in a diabetic boy are analyzed using techniques in computational geometry and in algebraic-geometric topology. We apply convolution and deblurring via diffusion equation from the side of computational geometry and knot theory, graph theory and singular homology from the side of algebraic-geometric topology. Our strategy consists in to represent the biomedical images using algebraic structures in such way that the peculiarities of the images are represented using algebraic complexities. With our strategy we obtain an automatic procedure for the analysis and the diagnostic based on biomedical images.

1 Introduction

The methods of Computer Vision can be applied to Biomedical Image Processing and the Computer Algebra Systems can be applied to Computer Vision [7]. In this paper we apply Computer Algebra to Biomedical Computer Vision. Specifically we analyze a case of rhinocerebral mucormycosis in a diabetic boy [8] using methods of Computer Vision such as convolution [7], deblurring [7], Tutte polynomials for graphs [1], Khovanov polynomials for knots [2] and Homology of simplicial complexes [3]. The advantages of Computer Algebra for Biomedical Computer Vision consist in that the combined symbolic and numerical strengths of the computer algebra systems are necessary to make automatic biomedical image analysis and processing [7]. We consider here a case of rhinosinusitis in a 14 year old diabetic boy originally presented at [8] which generates the consecutive biomedical images from computerized axial tomography depicted in Fig. 1. The aim of this paper is to present methods to describe the clinical progression from

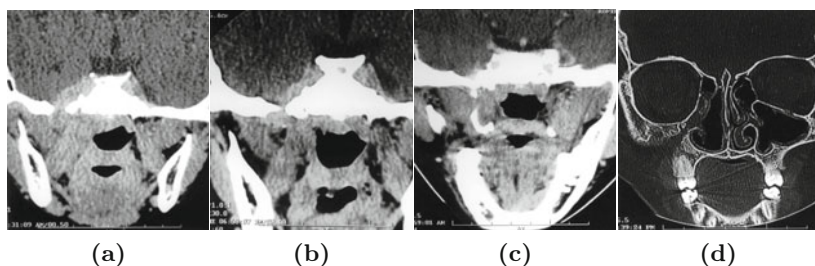


Fig. 1. Images taken from [8] and which will be analyzed in this work

the Fig. 1a, when the boy is sick to the Fig. 1d when the boy is healthy. We search for an automatic image processing that permits to discriminate between the biomedical images in Fig. 1. We claim that the automatic analysis of biomedical images will be important for the doctors as a tool for fast diagnostics.

2 Image Processing Using Convolution

The four images in Fig. 1 are analyzed in first instance using the method of convolution [7]. This method is automatically included in the Maple package ImageTools [4] and we start running the following maple algorithm:

```
> restart: with(ImageTools):
> imagefile := cat(kernelopts(datadir), "/help/ImageTools/
  chav1.jpg"):img := Read(imagefile):
> sharpened := Convolution(img, <<0,-1,0>|<-1,6,-1>|<0,-1,0>>,
  weight=20): View([img,sharpened]);
> imagefile2 := cat(kernelopts(datadir), "/help/ImageTools/
  chav2.jpg"): img2 := Read(imagefile2):
> sharpened2 := Convolution(img2, <<0,-1,0>|<-1,6,-1>|<0,-1,0>>,
  weight=20): View([img2,sharpened2]);
> imagefile3 := cat(kernelopts(datadir), "/help/ImageTools/
  chav3.jpg"): img3 := Read(imagefile3):
> sharpened3 := Convolution(img3, <<0,-1,0>|<-1,6,-1>|<0,-1,0>>,
  weight=20): View([img3,sharpened3]);
> imagefile4 := cat(kernelopts(datadir), "/help/ImageTools/
  chav4.jpg"): img4 := Read(imagefile4):
> sharpened4 := Convolution(img4, <<0,-1,0>|<-1,6,-1>|
  <0,-1,0>>,weight=20): View([img4,sharpened4]);
> View([img, img2, img3, img4]); View([sharpened, sharpened2,
  sharpened3, sharpened4]);
> View([img, img2, img3, img4, sharpened, sharpened2, sharpened3,
  sharpened4]);
```

When this algorithm is executed using the four images in Fig. 1 as the input, we obtain the output that is showed in Fig. 2. The four images in the bottom

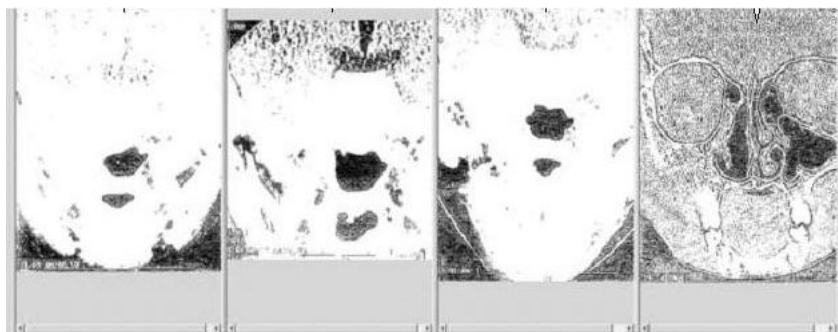


Fig. 2. Processing of the images in Fig. 1 using convolution

of Fig. 2 result from the four images in the top of Fig. 2 when the convolution algorithm previously presented is executed. We observe that the resulting images clearly show the difference between the three first ones images and the fourth image. In other words the convolution method separate the images with illness and the image with the healthy state.

3 Image Processing Using Deblurring via Diffusion Equation

The four images in Fig. 1 are analyzed now using deblurring via diffusion equation [7]. This method is implemented using an algorithm constructed with Mathematica [5]. The algorithm contains a part with symbolic computation and a part with numeric computation. The two parts, symbolic and numeric are necessary to obtain an automatic tool for the processing of the images in Fig. 1. The algorithm is as follows.

```

gD[im_List, nx_, ny_, s_] := Module[{x, y, kx, ky, mid, tmp},
g[x, s] = (1/(2*Pi*s^2)^(1/2))*Exp[-x^2/(2*s^2)];
kx = N[Table[Evaluate[D[g[x, s], {x, nx}], {x, -6*s, 6*s}]];
ky = If[nx == ny, kx,
N[Table[Evaluate[D[g[x, s] /. x -> y, {y, ny}],
{y, -6*s, 6*s}]]]; mid = Ceiling[Length[#1]/2]&;
tmp = Transpose[ListConvolve[{kx}, im, {{1, mid[kx]},
{1, mid[kx]}]]];
Transpose[ListConvolve[{ky}, tmp, {{1, mid[ky]},
{1, mid[ky]}]]];
deblur2d[im2d_, sigest_, order_, sig_] :=
Module[{expr}, DD2 = D[#1, {x, 2}] + D[#1, {y, 2}] &;
expr = Normal[Series[L[x, y, t], {t, 0, order}]] /.
Derivative[0, 0, n_][L_][x_, y_, t_] := Nest[DD2,
L[x, y, t], n]/. t -> (-sigest^2 + sig^2)/2;
Drop[expr, 1] /. Derivative[p_, q_, 0][L][x, y, t_] ->
HoldForm[gD[im2d, p, q, sig]]]

```

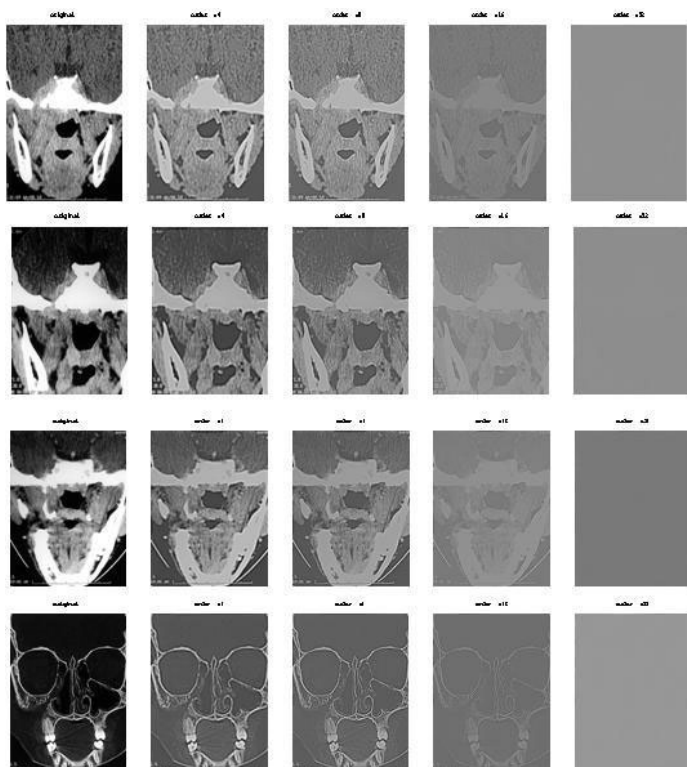


Fig. 3. Processing of the images in Fig. 1 using deblurring with diffusion equation

When this algorithm is executed using the four images in Fig. 1 as the input, we obtain the output that is showed in Fig. 3. We observe that the resulting images in Fig. 3 clearly show the difference between the three firsts rows with images and the fourth row. In other words the deblurring method separate the images with illness and the image with the healthy state.

4 Image Processing Using Geometric Topology

In this section we use methods in algebraic and geometric topologic with the aim to analyze the images in Fig. 1. In first instance we use graph theory and Tutte polynomials, in the second place we use knot theory and Khovanov polynomials and finally we use singular homology and Poincaré polynomials. With all these techniques we will give an algebraic characterization of the images in Fig. 1.

4.1 Image Processing Using Tutte Polynomials for Graphs

The four images in Fig. 1 are analyzed here using methods in geometric topology such as graph theory and Tutte polynomials. These methods are automatically included in the Maple package GraphTheory [4] and we start transforming the images in Fig. 1 into the graphs showed in Fig. 4.

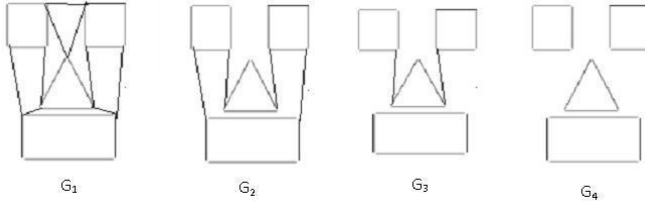


Fig. 4. Processing of the images in Fig. 1 using Graphs and Tutte polynomials

Then using the GraphTheory package for these graphs we obtain the following Tutte polynomials:

$$\begin{aligned}
 T(G_1, x, y) = & 68x + 68y + 473x^2 + 1487x^3 + 331y^2 + 736yx + 2892yx^2 + \\
 & 2894x^4 + 6166yx^3 + 6077y^2x^2 + 9150y^2x^3 + 8527yx^4 + 8385yx^5 + 6180yx^6 + \\
 & 3361y^3x + 6411y^3x^2 + 6822y^3x^3 + 8943y^2x^4 + 6121y^2x^5 + 3062y^2x^6 + \\
 & 3520yx^7 + 1568yx^8 + 3975x^5 + 2249y^2x + 4133x^6 + 3394x^7 + 702y^3 + \\
 & 2258x^8 + 867y^4 + 1231x^9 + 4028y^4x^2 + 4677y^3x^4 + 2209y^3x^5 + 1141y^2x^7 + \\
 & 312y^2x^8 + 540yx^9 + 29y^3x^8 + 57y^2x^9 + 23yx^{11} + 6y^5x^6 + 16y^4x^7 + 34y^6x^4 + \\
 & 2y^6x^5 + 20y^7x^3 + 10y^8x^2 + 4y^9x + 137yx^{10} + 2y^3x^9 + 5y^2x^{10} + 2yx^{12} + \\
 & y^4x^8 + x^{14} + 2983y^4x + 3063y^4x^3 + 748y^3x^6 + 706y^5 + 195x^{11} + 53x^{12} + \\
 & 1749y^3x + 406y^6 + 1683y^5x^2 + 720y^6x + 170y^7 + 10x^{13} + y^{10} + 1490y^4x^4 + \\
 & 489y^4x^5 + 932y^5x^3 + 494y^6x^2 + 210y^7x + 109y^4x^6 + 181y^3x^7 + 309y^5x^4 + \\
 & 63y^5x^5 + 188y^6x^3 + 96y^7x^2 + 40y^8x + 51y^8 + 10y^9 + 548x^{10}
 \end{aligned}$$

$$\begin{aligned}
 T(G_2, x, y) = & x + y + 8x^2 + 26x^3 + 4y^2 + 11yx + 43yx^2 + 46x^4 + \\
 & 81yx^3 + 63y^2x^2 + 65y^2x^3 + 89yx^4 + 81yx^5 + 81yx^6 + 25y^3x + 29y^3x^2 + \\
 & 11y^3x^3 + 45y^2x^4 + 45y^2x^5 + 39y^2x^6 + 77yx^7 + 65yx^8 + 54x^5 + 27y^2x + \\
 & 54x^6 + 6y^3 + 54x^7 + 53x^8 + 4y^4 + 49x^9 + 44x^9y + 40x^{10} + 27x^{11} + y^4x^2 + \\
 & 11y^3x^4 + 7y^3x^5 + 27y^2x^7 + 12y^2x^8 + 22yx^{10} + 14x^{12} + 5x^{13} + 7x^{11}y + \\
 & 8y^4x + y^4x^3 + 3y^3x^6 + 3x^9y^2 + y^5 + x^{12}y + x^{14}
 \end{aligned}$$

$$T(G_3, x, y) = x^2(y + x + x^2)(y + x + x^2 + x^3)^3$$

$$T(G_4, x, y) = (y + x + x^2)(y + x + x^2 + x^3)^3$$

As we see the algebraic complexity of the Tutte polynomials decrease when we pass from the image represented by G_1 to the image represented by

G_4 . This indicates that we have here algebraic discrimination of the images with illness (G_1, G_2, G_3) respect to the image with healthy state (G_4). Then the geometric topology and specifically the graph theory with its Tutte polynomial is able to generate automatic biomedical image processing for clinical applications.

4.2 Image Processing Using Khovanov Polynomials for Knots

The four images in Fig. 1 are analyzed here using methods in geometric topology such as knot theory and Khovanov homology. These methods are automatically included in the Mathematica package KnotTheory [6] and we start transforming the images in Fig. 1 into the knots showed in Fig. 5.

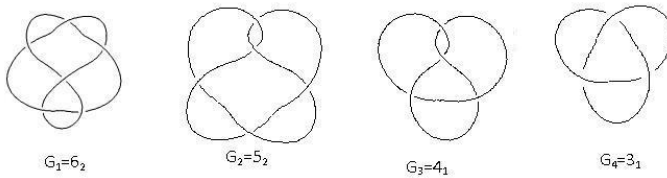


Fig. 5. Processing of the images in Fig. 1 using Knots and Khovanov polynomials

Then using the KnotTheorypackage for these knots we obtain the following Khovanov polynomials:

$$KH(6_2; q, t) = \frac{1}{q^3} + \frac{2}{q} + \frac{1}{q^{11} t^4} + \frac{1}{q^9 t^3} + \frac{1}{q^7 t^3} + \frac{1}{q^7 t^2} + \frac{1}{q^5 t^2} + \frac{1}{q^5 t} + \frac{1}{q^3 t} + \frac{t}{q} + q^3 t^2 \tag{1}$$

$$KH(5_2; q, t) = \frac{1}{q^3} + \frac{1}{q} + \frac{1}{q^{13} t^5} + \frac{1}{q^9 t^4} + \frac{1}{q^9 t^3} + \frac{1}{q^7 t^2} + \frac{1}{q^5 t^2} + \frac{1}{q^3 t} \tag{2}$$

$$KH(4_1; q, t) = \frac{1}{q} + q + \frac{1}{q^5 t^2} + \frac{1}{qt} + qt + q^5 t^2 \tag{3}$$

$$KH(3_1; q, t) = \frac{1}{q^3} + \frac{1}{q} + \frac{1}{q^9 t^3} + \frac{1}{q^5 t^2} \tag{4}$$

As we see the algebraic complexity of the Khovanov polynomials decrease when we pass from the image represented by G_1 (associated by the knot 6_2) to the image represented by G_4 (associated by the knot 3_1). This indicates that we have here algebraic discrimination of the images with illness (G_1, G_2, G_3) respect to the image with healthy state (G_4). Then the geometric topology and specifically the knot theory with its Khovanov homology is able to generate automatic biomedical image processing for clinical applications.

4.3 Image Processing Using Homology

The four images in Fig. 1 are analyzed here using methods in algebraic topology such as homology and the betti and the euler numbers. These methods are automatically included in the Maple package *Moise* [3] and we start transforming the images in Fig. 1 into the simplicial complexes showed in Fig. 6.

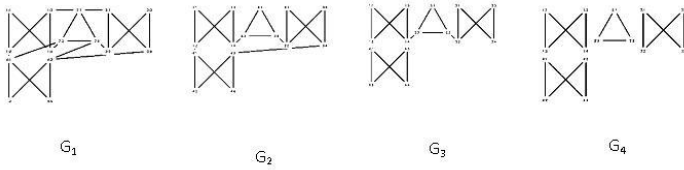


Fig. 6. Processing of the images in Fig. 1 using Homology for Simplicial Complexes

Then using the *Moise* package for these simplicial complexes we obtain the results that the Table 1 shows.

Table 1. Homological Analysis for the images of Fig. 6

Homology	Euler number	Poincare Polynomial
$Hom(G_1) = [[1, 10], [[], []]]$	$e(G_1) = -9$	$P(G_1, t) = t^3 + 15t^2 + 24t$
$Hom(G_2) = [[1, 5], [[], []]]$	$e(G_2) = -4$	$P(G_2, t) = t^3 + 15t^2 + 19t$
$Hom(G_3) = [[2, 4], [[], []]]$	$e(G_3) = -2$	$P(G_3, t) = t^3 + 15t^2 + 17t$
$Hom(G_4) = [[4, 4], [[], []]]$	$e(G_4) = 0$	$P(G_4, t) = t^3 + 15t^2 + 15t$

In Table 1 for every simplicial complex in Fig. 6 the homology, the euler number and the Poincare polynomial are showed. As we see the homology shows that the number of connected components grows when we pass from the image represented by \$G_1\$ (with only one connected component) to the image represented by \$G_4\$ (with four connected components). Reciprocally the homology shows that the number of holes decreases when we pass from the image represented by \$G_1\$ (with 10 holes) to the image represented by \$G_4\$ (with 4 holes). For other side, we observe that the euler number grows when we pass from the image represented by \$G_1\$ (euler number = -9) to the image represented by \$G_4\$ (euler number = 0). All these indicate that we have here algebraic discrimination of the images with illness (\$G_1, G_2, G_3\$) respect to the image with healthy state (\$G_4\$). Then the homology is able to generate automatic biomedical image processing for clinical applications.

5 Conclusions

An example of automatic analysis of biomedical images was presented. Some algebraic methods were used to transform the images in algebraic objects which were characterized and discriminated according with clinical and biomedical meanings. The main tool was computer algebra and specifically Maple and Mathematica were used jointly with complementary packages. The results in this work show that the computer algebra is useful for the automatic analysis and diagnostic in complex clinical situations. It is expected that the future parallel computer algebra systems will be very important for biomedical image processing including 3D computer vision and topological quantum computer vision.

References

1. Ellis-Monaghan, J., Moreno, C.: Graph polynomials and their applications I: The Tutte polynomial, arXiv:0803.3079 (2008), <http://xxx.lanl.gov/abs/0803.3079>
2. Turner, P.: Five Lectures on Khovanov Homology, arXiv:math/0606464 (2006), <http://xxx.lanl.gov/abs/math/0606464>
3. Moise: Maple Package, <http://www.maplesoft.com/applications/view.aspx?SID=370&L=F>
4. Maple, <http://www.maplesoft.com/>
5. Mathematica, <http://www.wolfram.com/>
6. KnotTheory: Mathematica Package, <http://www.math.toronto.edu/~drorbn/KAtlas/Manual/KnotTheoryManual.html>
7. Bart, M., Ter Haar Romeny, M.: Computer Vision and Mathematica. In: Computing and Visualization Science, Springer, Heidelberg (2002), <http://bmia.bmt.tue.nl/People/BRomeny/publications/ComputerVisionandMathematica.pdf>
8. Ana Chavez, P., Maritza Rahal, E.: Rhinocerebral mucormycosis in a diabetic child. Conservative treatment, Rev. chil. Infectol. 21(1) (2004), http://www.scielo.cl/scielo.php?pid=S0716-10182004000100008&script=sci_arttext

Applications

Adaptive B-Spline Model Based Probabilistic Active Contour for Weld Defect Detection in Radiographic Imaging

Nafaa Nacereddine¹, Latifa Hamami²,
Djemel Ziou³, and Aicha Baya Goumeidane¹

¹ LTSI, Centre de Recherche en Soudage et Contrôle, Chéraga, Alger, Algérie
e-mail: nafaa.nacereeddine@enp.edu.dz, ab_goumeidane@yahoo.fr

² Dépt. d'électronique, Ecole Nationale Polytechnique,
El-Harrach, Alger, Algérie
e-mail: latifa.hamami@enp.edu.dz

³ Dépt. d'informatique, Univ. de Sherbrooke,
Sherbrooke, Qc., Canada J1K 2R1
e-mail: djemel.ziou@usherbrooke.ca

Summary. This paper describes a probabilistic region-based deformable model using a new adaptive scheme for B-spline representation. The idea is to adapt the number of spline control points which are necessary to describe an object with complex shape. For this purpose, the curve segment length (CSL) is used as criterion. The proposed split and merge strategy on the spline model consists in: adding a new control point when CSL is greater than a certain splitting threshold so that the contour tracks all the concavities and, removing a control point when CSL is less to a certain merging threshold so that the contour aspect maintains its smoothness. Noise on synthetic and real weld radiographic images is assumed following Gaussian or Rayleigh distribution. The experiments carried out confirm the adequacy of this approach, especially in tracking pronounced concavities contained in images.

1 Introduction

In the radiographic testing (RT) of welds, the expert radiographer often works in extreme cases of the human visual system because of a low dimension of certain weld defects (e.g., a fissure can have a thickness lower than $200\ \mu\text{m}$), a bad contrast and a noised nature of the radiographic film, whereas perfect knowledge of the geometry of these weld defects is an important step which is essential to appreciate the quality of the weld [1]. In computer vision system dedicated to the evaluation of weld radiographic images, the segmentation consists in detecting the weld defect, considered as an object on a background. For many applications, in computer vision and imagery systems, number of

approaches have been developed to automatically recover the shape of objects from images. Active contours (or snakes or deformable model) and their conceptual descendants, since their introduction by Kass et al. [2], have been often proposed to deal with object contour estimation in several application contexts. The major interest of this approach is that it uses the knowledge that the object contour is simply connected, contrary to the traditional edge detectors [3]. Although many progresses [4, 5, 6, 7, 8, 9, 10, 11] have been brought to Kass et al. snake to deal with problems of sensitivity to initialization, inability to reparametrize during the deformation process, difficulties in progressing into boundary concavities and non-adaptiveness to topology changes, the improved versions still suffer from high noise and bad contrast in an image. This is caused by a strict utilization of information along the boundary without taking into account the information given by the regions inside and outside the contour, the reason for which this type of deformable models are called : contour-based active contours. On the other hand, region-based approach of active contours, used at first by Cohen et al. [12] and Ronfard [13], which is interested by both regions delimited by the contour, tries to solve problems of high noise, blurring, etc. in various applications. The region-based deformable models [3, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23] can be supported on statistical modeling of regions to segment [3, 17, 18, 23]. In this paper such modeling based active contour using an adaptive B-spline representation [24] is proposed.

2 Probabilistic Deformable Model

2.1 Probabilistic Image Model

Let $\mathbf{c} = \{c_0, \dots, c_{N-1}\}$ be the boundary of a connected image region R_1 of the plane and R_2 the set of points not included in R_1 . Denote z_i as the image gray-level observed at i -th pixel, $\mathbf{z} = \{z_i\}$ as the set of image gray levels, p_z as the gray-level density, and $\psi_z = \{\psi_1, \psi_2\}$ as the density parameters (i.e., $p_z(z_i) = p_z(z_i|\psi_1)$ for $i \in R_1$ and $p_z(z_i) = p_z(z_i|\psi_2)$ for $i \in R_2$). The simplest possible region-based model is characterized by the two following hypotheses: conditional independence (given the region boundary, all the pixels are independent); and region homogeneity (the probability distribution of each pixel only depends on whether it belongs to the regions R_1 or R_2). Thus, the likelihood function can be written as [17]

$$p_{\mathbf{z}|\mathbf{c}}(\mathbf{z}|\mathbf{c}, \psi_z) = \prod_{i \in R_1} p_z(z_i|\psi_1) \prod_{i \in R_2} p_z(z_i|\psi_2) \quad (1)$$

2.2 Bayesian Approach for Contour Estimation

Here, the contours are built as vectors of a subspace $\mathcal{R}(\mathbf{B}_k)$ where \mathbf{B}_k is the cubic B-spline basis and k the number of control points. In other words, $\mathbf{c} = \mathbf{B}_k \alpha_k$ where $\alpha_k = \{\alpha_1, \dots, \alpha_{k-1}\}$ is a $(k \times 2)$ vector containing the coordinates

of the control points. The B-spline formulation of the active contour does not allow only a local control of the curve by handling individually each control point, but proposes also a compact representation which possesses two main advantages: its construction requires only few parameters (3 in case of cubic B-spline) and the regularity is intrinsic to its structure [25]. We assume that contours $\mathbf{c}(k) = \mathbf{c}(k, \alpha_k)$ are random vectors with probability density function given by $p_c(\mathbf{c}(k)) = p_k(k|\psi_c)$, where ψ_c denotes a parameter vector of p_k . Hence, the maximum a posteriori (MAP) estimate of the pair (\mathbf{c}, k) is

$$(\hat{c}, \hat{k}) = \arg \max_{k, \mathbf{c} \in \mathcal{R}(\mathbf{B}_k)} \{p_{\mathbf{z}|c}(\mathbf{z}|c, \psi_z)p_k(k|\psi_c)\} \quad (2)$$

2.3 Case with Fixed k

The MAP estimate where k is known is simply the contour estimate in the sense of maximum likelihood (ML). Moreover, if the parameter vector ψ_z is unknown, the ML estimate of (c, ψ_z) is determined as

$$(\hat{c}, \hat{\psi}_z) = \arg \max_{c \in \mathcal{R}(\mathbf{B}_k), \psi_z} \{L_{\mathbf{z}|c}(\mathbf{z}|c, \psi_z)\} \quad (3)$$

where $L_{\mathbf{z}|c}(\mathbf{z}|c, \psi_z) \equiv \log\{p_{\mathbf{z}|c}(\mathbf{z}|c, \psi_z)\}$ is the log-likelihood function. The log function is included, since it does not affect the location of the maximum and allows some formal simplification. Since solving (3) simultaneously with respect to \mathbf{c} and ψ_z would be computationally very difficult, we settle for a suboptimal solution given by iterative schemes of the type [18]

$$\hat{c}^{(t+1)} = \arg \max_{c \in \mathcal{R}(\mathbf{B}_k)} \{L_{\mathbf{z}|c}(\mathbf{z}|c, \hat{\psi}_z^{(t)})\} \quad (4)$$

$$\hat{\psi}_z^{(t+1)} = \arg \max_{\psi_z} \{L_{\mathbf{z}|c}(\mathbf{z}|\hat{c}^{(t+1)}, \psi_z)\} \quad (5)$$

where $\hat{c}^{(t)}$ and $\hat{\psi}_z^{(t)}$ are the estimates of \mathbf{c} and ψ_z at iteration t , respectively. The computation of $(\hat{c}, \hat{\psi}_z)$ can be summarized by the Algorithm 1.

2.4 Case with Adaptive k

In this case, the displacement Δc is computed along orthogonal lines. This choice is justified by the fact that for the active contour external and internal energies, it can be shown that if the optimal deformation is small, then it needs be only normal to the template [26], i.e. orthogonal to the tangent vector dc/dt where s is the curvilinear abscissa of the contour. Also, with this configuration, by using both senses of the orthogonal line instead of the 8-neighborhood positions, the space search is reduced from 1 to 1/4. In addition, we present in this section our contribution in this paper, namely a resampling process based on split and merge operations in order to allow the contour represented by B-spline to adapt the number of control points

which are necessary to describe accurately the object. The length of each curve segment is used as criterion to adapt the number of control points. In fact, if the length of a curve segment exceeds a certain value d_s , computed as the product of the average length value l_a of all curve segments by a certain split threshold $\delta_s (> 1)$, i.e. $d_s = \delta_s \times l_a$, the curve segment is subdivided by adding a new control point. Conversely, if the length of a curve segment is under d_m , where $d_m = \delta_m \times l_a$ with $\delta_m < 1$ is the merging threshold, then the last control point of the segment concerned is eliminated. This process is repeated until stabilization in sense of $L_{\mathbf{z}|c}$. In this case, $(\hat{c}, \hat{\psi}_z)$ can be estimated by the iterative Algorithm 2.

Algorithm 1

Initialization : set $c^{(0)}$, $\psi_z^{(0)}$ and ϵ

Do

$$\text{Step 1: } \Delta\hat{c}^{(t)} = \arg \max_{u \in \Omega(\hat{c}^{(t)})} \{L_{\mathbf{z}|c}(\mathbf{z}|\hat{c}^{(t)} + u, \hat{\psi}_z^{(t)})\}$$

where $\Omega(c) \subset \mathcal{C}^N$ is the set of points defining the 8 neighbors of the control point displacement

$$\text{Step 2: } \hat{c}^{(t+1)} = \hat{c}^{(t)} + \mathbf{B}_k \Delta\hat{c}^{(t)}$$

$$\text{Step 3: } \hat{\psi}_z^{(t+1)} = \arg \max_{\psi_z} \{L_{\mathbf{z}|c}(\mathbf{z}|\hat{c}^{(t+1)}, \psi_z)\}$$

$$\text{Step 4: } \Delta L = L_{\mathbf{z}|c}^{(t+1)} - L_{\mathbf{z}|c}^{(t)}$$

While $|\Delta L| \geq \epsilon$

Algorithm 2

Initialization : set $c^{(0)}$, $\psi_z^{(0)}$, ϵ_1 , ϵ_2 , δ_s , δ_m

Do (loop 2)

Do (loop 1)

$$\text{Step 1: } \Delta\hat{c}^{(t_1)} = \arg \max_{u \in \Omega(\hat{c}^{(t_1)})} \{L_{\mathbf{z}|c}(\mathbf{z}|\hat{c}^{(t_1)} + u, \hat{\psi}_z^{(t_1)})\}$$

where $\Omega(c) \subset \mathcal{C}^N$ is the set of points defining the orthogonal displacement to the contour \mathbf{c}

$$\text{Step 2: } \hat{c}^{(t_1+1)} = \hat{c}^{(t_1)} + \mathbf{B}_k \Delta\hat{c}^{(t_1)}$$

$$\text{Step 3: } \hat{\psi}_z^{(t_1+1)} = \arg \max_{\psi_z} \{L_{\mathbf{z}|c}(\mathbf{z}|\hat{c}^{(t_1+1)}, \psi_z)\}$$

$$\text{Step 4: } \Delta L_1 = L_{\mathbf{z}|c}^{(t_1+1)} - L_{\mathbf{z}|c}^{(t_1)}$$

While $|\Delta L_1| \geq \epsilon_1$

$$L_{\mathbf{z}|c}^{(t_2)} = L_{\mathbf{z}|c}^{(t_1+1)}$$

$$\text{Step 5: } \hat{c}^{(t_2)} = \text{Split}(\hat{c}^{(t_1+1)}, \delta_s)$$

$$\text{Step 6: } \hat{c}^{(t_2)} = \text{Merge}(\hat{c}^{(t_2)}, \delta_m)$$

Go to loop 1

Deduce $L_{\mathbf{z}|c}^{(t_2+1)}$

$$\text{Step 7: } \Delta L_2 = L_{\mathbf{z}|c}^{(t_2+1)} - L_{\mathbf{z}|c}^{(t_2)}$$

While $|\Delta L_2| \geq \epsilon_2$

3 Experimental Results

The expression (5) depends on the particular structure of p_z . For example, for Gaussian densities with mean μ and variance σ^2 , estimates of $\psi_z \equiv \{\mu_i, \sigma_i^2\}$, for $i = 1, 2$, are given by the mean and variance within the regions of object and background, respectively; whilst for Rayleigh distribution which is given by $p(z/\phi) = \frac{z}{\phi} \exp\left[-\frac{z^2}{2\phi}\right]$; $z \geq 0$, $\psi_z \equiv \{\phi_1, \phi_2\}$, where ϕ_1 and ϕ_2 are the Rayleigh parameters for the object and background regions, respectively. As illustrated in Fig. 1, the implementation results of Algorithm 1 on synthetic images, corrupted by Gaussian-distributed (IG,JG) and Rayleigh-distributed (IR,JR) noises, show successful contour estimates from various initializations, proving the robustness of this region-based deformable model approach with respect to contour initialization. Indeed, the initial contour has just to cross the real contour or to be completely inside or surrounding the real contour, to guarantee the convergence. Moreover, although the noised nature of the considered images, the object contours are correctly estimated. Nevertheless, let us note that the tested images do not contain objects with complex shapes and for these images, between 9 and 12 control points are sufficient to adapt the B-spline curves to the apparent contours.

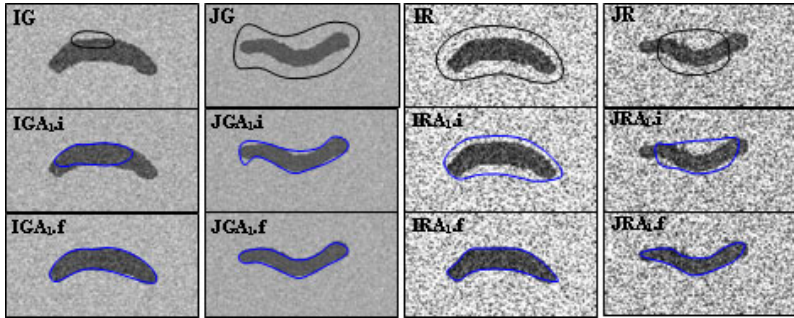


Fig. 1. Estimated contours by Algorithm 1 on synthetic images corrupted by Gaussian- and Rayleigh-distributed noises. Intermediate (i) and final (f) contours.

Also, we have implemented Algorithm 1 (Algo.1) and Algorithm 2 (Algo.2) on a synthetic image with complex shape, corrupted by Gaussian noise with parameters : $\{\mu_1, \sigma_1\} = \{130, 10\}$, $\{\mu_2, \sigma_2\} = \{70, 10\}$. It seems from results given in Fig. 2 that Algo.1 is unable to adapt the estimated contour on the object boundary for both initialization schemes which use 12 control points. In fact, during the contour evolution, the control points displacements, enacted by the maximization of the likelihood function and constrained in the 8-neighborhood, can not track the concavities really contained in the object. This situation leads to control points "pile-up" in some contour parts and a lack of control points in the others parts. This is why, the contour

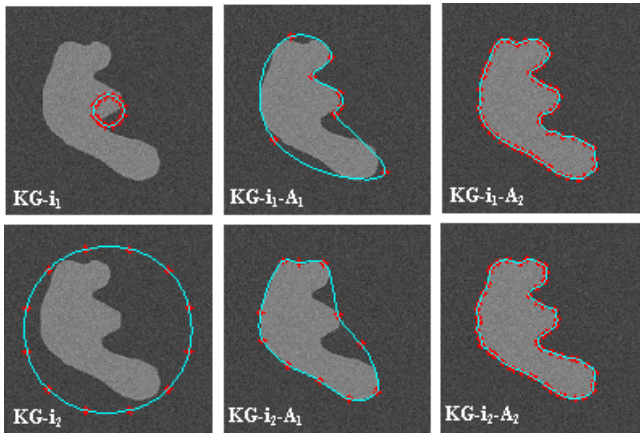


Fig. 2. Gaussian noise corrupted synthetic image, initializations (i_1, i_2) and estimated contours using Algo.1 and Algo.2

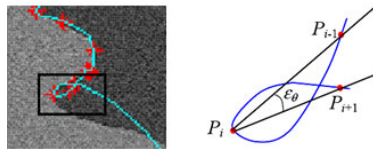


Fig. 3. Loop in a cubic B-spline

estimation scheme given by Algo.2 can remedy this problem where control points are added so that the contour tracks the concavities and are removed so that the contour aspect maintains its smoothness, as shown in Fig. 2. The relative errors between the original parameters of the object and background and those estimated confirm the good comportment of Algo.2 in the contour estimation of complex shape object, where $\{\Delta\mu_1, \Delta\sigma_1\}$ and $\{\Delta\mu_2, \Delta\sigma_2\}$ are less respectively, to $\{0.5, 17\}$ and $\{0.2, 3\}$ (in %) for both initializations. It is not rare to find loops in a cubic B-spline, as illustrated in Fig. 3. For their elimination, we have carried out a test which consists to remove a control point if it composes with its neighbors an angle less than a certain limit angle ϵ_θ .

Also, it is convenient to compare the proposed algorithms on a real weld defect image shown in Fig. 4, where a reference binary image has been extracted manually by an expert in radiography. To show the importance of the noise distribution law, we consider this image corrupted by two noise models, namely those of Gauss and Rayleigh. Then the parameters of ground truth for both distributions are computed *a priori*. The same initial contour with the same number of control points is used for both laws and both algorithms. It appears through the results given in Fig. 4 and confirmed by the parameter

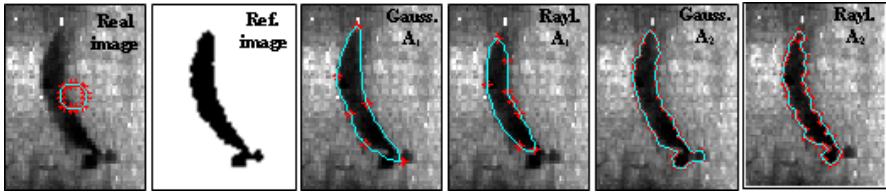


Fig. 4. Real weld defect image, binary reference image, estimated contours by Algo.1 and Algo.2 using Gaussian and Rayleigh models

Table 1. Real image inner and outer region parameters for the ground truth and the estimated contour

Ground truth		Algo. 1		Algo. 2	
Gaussian model					
μ_1, σ_1	μ_2, σ_2	$\hat{\mu}_1, \hat{\sigma}_1$	$\hat{\mu}_2, \hat{\sigma}_2$	$\hat{\mu}_1, \hat{\sigma}_1$	$\hat{\mu}_2, \hat{\sigma}_2$
27.47 19.05	136.08 45.63	31.06 19.77	135.90 46.01	27.92 18.60	136.23 45.31
Rayleigh model					
ϕ_1	ϕ_2	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_1$	$\hat{\phi}_2$
23.92	101.49	20.94	100.45	22.20	101.08

estimation values provided in Table 1, that Algo.2 is, in one hand, more accurate in the recovering of the real defect boundary and, in the other hand, the best results are obtained with the noise Gaussian distribution. What is more realistic because this type of images (industrial or medical radiographic image) follow, in general, Gaussian distribution and that is due mainly to the differential absorption principle which governs the formation process of such images. The various parameters used for both algorithms are summarized as follows: $\epsilon = 10^{-6}, \epsilon_1 = 10^{-6}, \epsilon_2 = 10^{-6}, \delta_s = 1.2, \delta_m = 0.6, \epsilon_\theta = \pi/45$.

4 Conclusion

In this work, the first algorithm describes an approach of region-based active contour based on the likelihood maximization using cubic B-spline with a fixed number of control points. This B-spline configuration can stop the contour evolution before it reaches global maximum likelihood, especially for complex shape objects. The parametrization of B-spline so that it modifies the number of control according a split and merge strategy can overcome this situation. The outstanding performance of this approach is confirmed by experiments on synthetic and real weld defect images.

References

1. Schwartz, C.: Automatic evaluation of welded joints using image processing on radiographs. In: Conf. Proceed. of American Inst. of Physics, vol. 657, pp. 689–694 (2003)
2. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *Int. J. Computing and Vision* 1(4), 321–331 (1988)
3. Chesnaud, C., Réfrégier, P., Boulet, W.: Statistical region snake-based segmentation adapted to different physical noise models. *IEEE Trans. on PAMI* 21(11), 1145–1157 (1999)
4. Cohen, L.D.: On active contour models and balloons. *CVGIF: Image Understanding* 53, 211–218 (1991)
5. Cohen, L.D., Cohen, I.: Finite-element methods for active contour models and balloons for 2-D and 3-D images. *IEEE Trans. on PAMI* 15(11), 1131–1147 (1993)
6. Leroy, B., Herlin, I., Cohen, L.D.: Multi-resolution algorithms for active contour models. In: 12th Int. Conf. Analysis and Optimization of Systems, pp. 58–65 (1996)
7. Xu, C., Prince, J.: Snakes, shapes, and gradient vector flow. *IEEE Trans. Image Processing* 7(3), 359–369 (1998)
8. Xu, C., Prince, J.: Generalized gradient vector flow: External forces for active contours. *Signal Processing* 71, 131–139 (1998)
9. Malladi, R., Sethian, J., Vemuri, B.: Shape modeling with front propagation: A level set approach. *IEEE Trans PAMI* 17, 158–175 (1995)
10. Caselles, V., Catta, F., Coll, T., Dibos, F.: A geometric model for active contours in image processing. *Num. Math.* 66, 1–31 (1993)
11. Goumeidane, A.B., Khamadja, M., Nacereddine, N., Mekhalfa, F.: Parametric active aontour for weld defect boundary extraction in radiographic testing. In: SPIE conf. on Quality Control by Artificial vision, Le Creusot, France (2007)
12. Cohen, L., Bardinet, E., Ayache, N.: Surface reconstruction using active contour models. In: SPIE conf. on Geometric Methods in Computer Vision, San Diego, CA (1993)
13. Ronfard, R.: Region-based strategies for active contour models. *Int. Journal of Computer Vision* 13(2), 229–251 (1994)
14. Zhu, S., Yuille, A.: Region competition: unifying snakes, region growing, and bayes/MDL for multiband image segmentation. *IEEE Trans. on PAMI* 18, 884–900 (1996)
15. Germain, O., Réfrégier, P.: Optimal snake-based segmentation of a random luminance target on a spatially disjoint background. *Optics Letters* 21(22), 1845–1848 (1996)
16. Yezzi, A., Tsai, A., Willsky, A.: A statistical approach to snakes for bimodal and trimodal imagery. In: Int. Conf. on Image Processing, Kobe, pp. 898–903 (1999)
17. Dias, J.M.B.: Adaptive bayesian contour estimation: A vector space representation approach. In: Hancock, E.R., Pelillo, M. (eds.) *EMMCVPR 1999*. LNCS, vol. 1654, pp. 157–172. Springer, Heidelberg (1999)
18. Figueiredo, M.A.T., Leitao, J.M.N., Jain, A.K.: Unsupervised contour representation and estimation using B-splines and a minimum description length criterion. *IEEE Trans. on Image Processing* 9(6), 1075–1087 (2000)

19. Paragios, N., Deriche, R.: Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Trans. on PAMI* 22, 266–280 (2000)
20. Chan, T., Vese, L.: Active contours without edges. *IEEE Trans. on Image Processing* 10(2), 266–277 (2001)
21. Jehan-Besson, S., Barlaud, M., Aubert, G.: DREAM2S: Deformable regions driven by an Eulerian accurate minimization method for image and video segmentation. *Int. Journal of Computer Vision* 53(1), 45–70 (2003)
22. Herbulot, A., Jehan-Besson, S., Duffner, S., Barlaud, M., Aubert, G.: Segmentation of vectorial image features using shape gradients and information measures. *Journal of Mathematical Imaging and Vision* 25(3), 365–386 (2006)
23. Nacereddine, N., Hamami, L., Ziou, D., Tridi, M.: Probabilistic deformable model for weld defect contour estimation in radiography. *Machine Graphics & Vision* 15(3/4), 547–556 (2006)
24. Dierchx, P.: Curve and surface fitting with splines. Oxford Univ. Press, Oxford (1993)
25. Precioso, F., Barlaud, M., Blu, T., Unser, M.: Robust real-time segmentation of images and videos using a smooth-spline snake-based algorithm. *IEEE Trans. on Image Processing* 14(7), 910–924 (2005)
26. Tagare, H.: Deformable 2-D template matching using orthogonal curves. *IEEE Trans. on Medical Imaging* 16(1), 108–117 (1997)

FONN-Based Affine-Invariant Image Recognition

Bartłomiej Stasiak

Institute of Information Technology, Technical University of Łódź,
ul. Wólczajska 215, 93-005 Łódź, Poland
e-mail: basta@ics.p.lodz.pl

Summary. In this paper a fast orthogonal neural network (FONN) is used to construct an image classifier invariant to basic affine transformations (rotation, translation, scaling). The shift-invariance property of the Fourier amplitude spectrum in conjunction with the log-polar transform is applied for this purpose. Two image databases are built and used for testing the proposed classifier.

1 Introduction

Affine transformations of images, resulting from the natural variability of acquisition circumstances, are one of the main sources of ambiguities and difficulties in the field of image recognition (IR). Determining the position, orientation, size and other parameters of an object in a scene is an important preliminary stage (*image registration*) in many IR systems, which influences the reliability of classification and the recognition rate. This stage is in turn dependent on the segmentation of the image which should generally enable to find the objects of interest. Considering the fact that successful segmentation and analysis of complex scenes is still a problematic task, it is worth considering a question: could a classifier itself be invariant to affine transformations of unregistered images?

We are interested in raw image classification, which may seem difficult due to high dimensionality and redundancy of the input data, but which at the same time eliminates the risk of erroneous image segmentation and registration. We use analogy to shift invariance property of Fourier amplitude spectrum [3][1] to obtain a neural network invariant to input image translation. Additionally, Fourier-Mellin transform is used to provide the rotation and scaling invariance. In contrast to typical approaches in which Fourier amplitude spectrum of an image is computed and some of its coefficients then form a fixed feature vector for a classifier [4][2], we use fast orthogonal neural network (FONN) [7] to compute the spectrum in an adaptable way. The

whole classification system consists of a single, multilayer neural network in which the input layers provide the shift invariance property, as their sparse connection scheme is based on fast algorithm of Fourier transform, and the single output layer computes and assigns the final class labels. In this way we do not have to determine which spectral coefficients are important and should be included in the feature vector. Moreover, although the FONN part of the system may learn the exact amplitude spectrum values, it may also learn other transforms, possibly preserving the phase information. This information, rejected in a typical approach based on amplitude spectrum only, may be used to further enhance classification results.

2 The Structure of the FONN Classifier

The distinctive feature of fast orthogonal neural networks is their close relation to fast algorithms of orthogonal transforms (e.g. cosine or Fourier transforms). This relation is expressed in the sparse architecture of neural connections reflecting the factorization scheme of the orthogonal transform matrix [5][7]. The flow-graph of the fast algorithm is directly transformed into the neural network structure and its basic operations become the basic operation orthogonal neurons (BOONs) with two inputs and two outputs (Fig. 1) [7].

Naturally, considering the recursive nature of the fast algorithm construction it is possible to obtain a classifier for input of size $N \times N$ for any N being a power of 2. The number of neurons in the last layer corresponds to the number of classes. The connections denoted with (*) are simple one-to-one

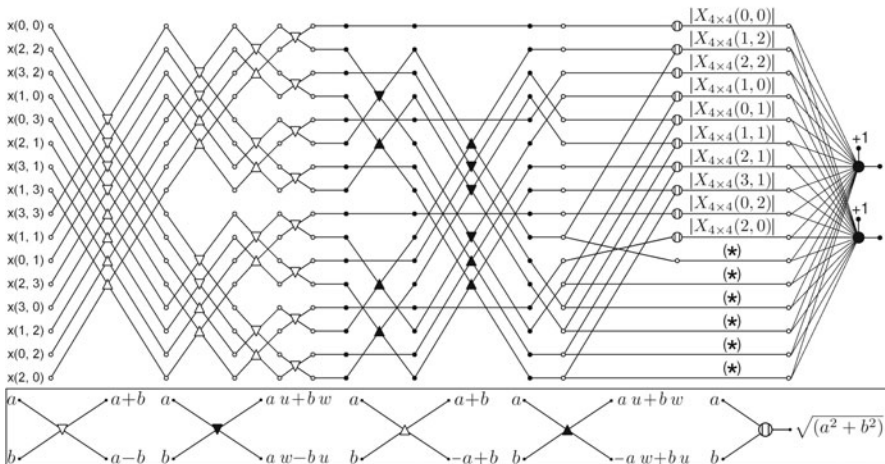


Fig. 1. The complete neural architecture for input images of size 4×4 and two target classes

connections enhancing the recognition in the test cases in which the phase information is also important [7].

2.1 Invariance to Affine Transformations

Fourier amplitude spectrum is inherently invariant to translation of the input image. The invariance to rotation and scaling may be obtained with the Fourier-Mellin transform [1][6]. For digital images, the computations may be split into two steps – first, transforming rotation and scaling into translations, and – second, computing the Fourier amplitude spectrum [3].

The first step is realized by changing the coordinate system from cartesian to log-polar [3][1]. The difficulty, which cannot be neglected here is the necessity of eliminating the image translation first because shifting the original image introduces huge differences in the log-polar representation [6]. The time-domain methods of obtaining the translation invariance such as the center of gravity, shape and contour descriptors [1][2] are substantially dependent on the quality of the segmentation process, which may not yield satisfactory results for real-world images. Therefore, in our approach we compute the Fourier amplitude spectrum, which is invariant to translation while preserving rotation and scaling (with the inverse of the scale factor), as the first step of the processing. The spectrum is then subjected to log-polar transform (LPT) which yields the FFT-LPT representation (Fig. 2) [3].

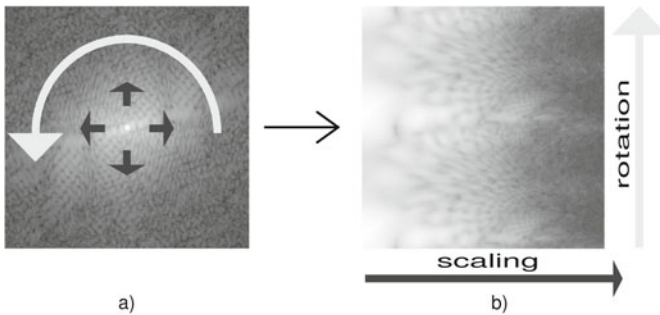


Fig. 2. a) Image amplitude spectrum; b) Its log-polar representation (FFT-LPT)

The log-polar transform was realized via inverse mapping [2], in which for every pixel of the resulting image with coordinates (ρ, φ) the coordinates (x, y) of the pixel in the original image were computed according to the relation:

$$[x, y] = \left[-r \cos \left(\frac{2\pi}{h-1} \varphi \right) + \frac{w_{or}}{2}, -r \sin \left(\frac{2\pi}{h-1} \varphi \right) + \frac{h_{or}}{2} \right], \quad (1)$$

where the size of the original and the resulting image is $(w_{or} \times h_{or})$ and $(w \times h)$, respectively; $r = \exp((\rho - c)/a) - b$, $\varphi = 0, 1, \dots, h - 1$, $\rho = 0, 1, \dots, w - 1$, and the parameters a and c are defined as:

$$a = \frac{w - 1}{r_{max} - r_{min}}, \quad c = \frac{(1 - w) \cdot (r_{min} + b)}{r_{max} - r_{min}}. \quad (2)$$

The parameters r_{min} and r_{max} decide on the area of the original image, which is subjected to the log-polar mapping. Their values, after some initial tests were set to: $r_{min} = 2$; $r_{max} = w_{or}/2 - 1 = h_{or}/2 - 1 = 63$.

Parameter b influences the shape of the resulting logarithmic curve and, hence, the size of the log-polar representation of various regions of the original image. However, as the goal of the log transform is to change scaling into translation, let us consider the image scaled with factor s . We would obtain:

$$s \cdot r = s \cdot e^{\frac{\rho - c}{a}} - s \cdot b = e^{\frac{\rho + a \ln(s) - c}{a}} - s \cdot b. \quad (3)$$

We can observe, therefore, that the parameter b should be equal to 0. Only then the scaling with factor s is equivalent to shifting the log-polar representation along the coordinate ρ by a value of $a \ln(s)$.

There are two ways to obtain complete rotation, translation and scaling (RTS) invariance with the LPT of the amplitude spectrum. We can simply compute the Fourier amplitude spectrum again (FFT-LPT-FFT) and perform classification with any classifier (e.g. standard K-nearest neighbors method, KNN) or apply the FONN classifier on the FFT-LPT representation for the same purpose. In the next section we will compare these two methods showing the advantages of the approach based on FONN with respect to KNN classifier.

3 Experimental Validation

Two separate databases were constructed for performing all the tests. In the first one, the **RTS3** database (Fig. 3), the objects were manually subjected to affine transformations and put on an artificial background. The second one, the **STaR** database, consists of photographs of real objects taken with varying perspective and lighting conditions.

The **RTS3** database contains images of 2 objects. Each object is used to construct two classes: in one class it is rotated by a random angle from the range $-30^\circ, +30^\circ$ or $(150^\circ, 210^\circ)$, which means in practice "roughly vertical position"; in the other class the angles are taken from either of the ranges $(60^\circ, 120^\circ)$ or $(240^\circ, 300^\circ)$, which yields "roughly horizontal position". Besides, the images within each class are varied by scale coefficient from the range $(0.9, 1.5)$ and position. All the objects are presented on background obtained by cutting a square fragment from a different image (classic *Fishing boat* image).

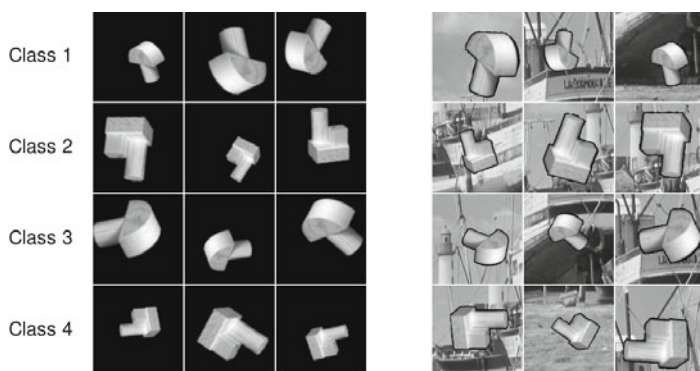


Fig. 3. The class representatives (three examples from each class) in the database **RTS3** on uniform background (left) and in their final form (right)

The database was generated several times with different number of training images to assess the generalization properties of the classifier. One hundred testing images were used in all cases and the validation dataset size was set to one-fourth of the training set (rounded up to the nearest integer). For every input image of size 128×128 its Fourier amplitude spectrum of the same size was computed and it was transformed to the log-polar representation of size 32×32 . The resulting 1024-element input vectors were normalized and the constant component was removed.

The results of four different classification procedures have been presented in Fig. 4. These are: KNN classifier performing recognition of raw images (black), of their FFT-LPT representation (dark gray) and of their FFT-LPT-FFT representation (white) and the FONN classifier using the FFT-LPT representation (light gray).

The database **RTS3** is constructed in such a way that a shift of the FFT-LPT representation along the vertical axis, corresponding to rotation of the spectrum and - at the same time - of the original image, cannot be neglected. In this case, only operating on FFT-LPT representation enables to perform correct recognition, which is shown by the KNN (FFT-LPT) and FONN classifiers. However, the important difference between them is that the FONN does provide the shift invariance, due to its structural analogy to the algorithm for Fourier amplitude spectrum computation, which does not exclude using also the phase information. In contrast, the KNN is a statistical method, which can "simulate" shift invariance if given sufficient training data, but it is basically unaware of the underlying model of affine transformations. The obtained results seem to confirm this reasoning.

In the second database (**STaR**) the backgrounds were much more homogeneous, but the number of classes was increased to ten. The recognized objects were small hand tools (one tool = one class). The photo of every object was taken ten times on three different backgrounds, so that the whole database

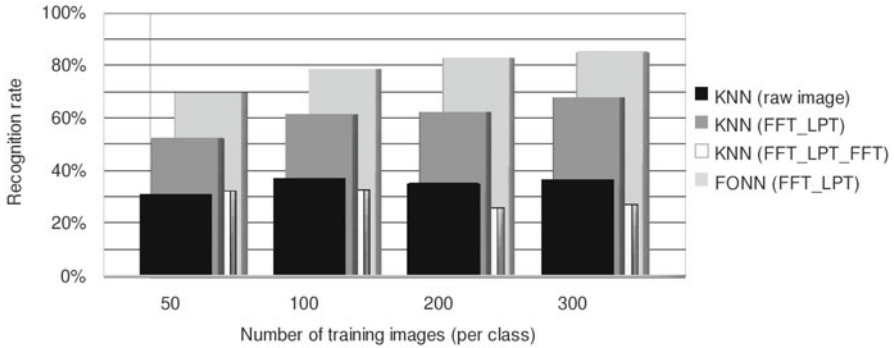


Fig. 4. Classification results (RTS3 database)

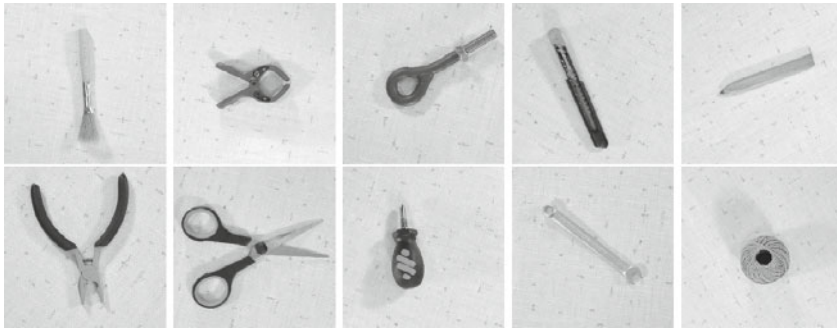


Fig. 5. Examples of correctly recognized images from the STaR-light set

consisted of 10×10 objects \times 3 backgrounds = 300 images. The position, orientation (planar rotation angle from the range $0^\circ - 360^\circ$) and distance from the camera (from the range 100% – 150%) were randomly set for every image. Fifty images (5 images per class) were chosen to form the test set and the remaining 250 were used as the training set.

Additionally, in order to determine the robustness of the network to greater variations of image acquisition conditions, two other sets of images were constructed, containing 50 images each. In the first one (the **STaR-light** set, Fig. 5) varied lighting conditions were used and in the second one (the **STaR-light30** set, Fig. 6) also the angle between the image plane and the lens plane was changed from 0° to 30° .

The neural classifier from Fig. 1, without the connections denoted by (*) was applied for classification tests and Fourier amplitude spectrum in the log-polar representation of size 64×64 (FFT-LPT) was used as input data.

Three experiments have been performed on the basis of the test sets composed of, respectively: 50 images¹ from the **STaR** database, 50 images from

¹ The remaining 250 images formed the training set in all the three experiments.

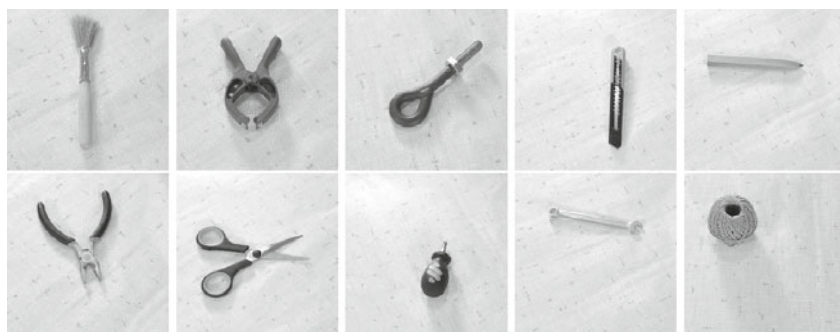


Fig. 6. Examples of correctly recognized images from the STaR-light30 set

Table 1. Confusion matrices for the STaR-light (left) and STaR-light + STaR-light30 (right) databases (each value shown is a percentage of correct recognition)

	brush	clamp	hook	knife	pencil	pliers	scissors	screwdriver	spanner	string	brush	clamp	hook	knife	pencil	pliers	scissors	screwdriver	spanner	string
brush	28	-	-	72	-	-	-	-	-	-	15	-	-	66	9	-	-	-	10	-
clamp	-	100	-	-	-	-	-	-	-	-	-	70	-	-	-	10	20	-	-	-
hook	8	-	82	-	10	-	-	-	-	-	5	-	60	9	-	18	-	8	-	-
knife	-	-	-	100	-	-	-	-	-	-	-	-	-	100	-	-	-	-	-	-
pencil	36	-	-	-	64	-	-	-	-	-	18	-	10	30	32	1	-	-	9	-
pliers	-	-	-	-	-	100	-	-	-	-	-	9	-	-	-	91	-	-	-	-
scissors	-	-	-	-	-	-	100	-	-	-	-	-	7	-	-	2	88	-	3	-
screwdriver	-	-	-	-	-	-	20	80	-	-	-	-	-	-	-	19	80	1	-	-
spanner	-	-	18	40	-	-	-	-	42	-	-	-	26	19	3	-	-	2	50	-
string	-	-	-	-	-	-	-	-	-	100	2	-	9	-	1	-	18	21	-	49

the STaR-light set, and 100 images from the combined STaR-light and STaR-light30 sets. The FONN classifier has reached the recognition rate of 100%, 79.6%, 63.5%, respectively. The confusion matrices for the second and the third experiment are presented in Table 1.

As a comparison, the recognition rate of the KNN classifier was below 60% even for the set from the first experiment. This fact, probably resulting from the big size of the input data (64 × 64) and low number of training images (25 objects per class), emphasizes good generalization properties of the FONN.

Increasing the rate of distortions resulting from lighting conditions and perspective change leads to quite significant decrease of the recognition rate. The analysis of the confusion matrix shows that the biggest problems occur in the case of three objects with elongated shape and low contrast with the background (brush, spanner, pencil), which are – subjectively – quite similar.

It should be noted, however, that all those additional distortions are not included in the training set. Adding the images typical for the **STaR-light** and **STaR-light30** sets into the training set would probably lead to better recognition results.

4 Conclusion

The presented classifier based on a fast orthogonal neural network has proven its effectiveness in the task of recognition of images under basic affine transformations (rotation, translation, scaling). The flexibility of the proposed approach, resulting from the adaptation potential of a neural network, is shown in a task involving rotation as a partially discriminative factor (the **RTS3** database).

Finally, it is worth mentioning that although the **STaR** database seems to be an easier classification problem in comparison to the **RTS3** set, mainly due to the possibility of applying potentially successful segmentation techniques and manually choosing appropriate features for discriminating the objects, the simplicity and generality of the proposed approach is its major advantage. The proposed classifier may be successfully applied in practice to recognize objects without the necessity of finding their localization, orientation and size and without additional preprocessing stages such as segmentation.

References

1. Derrode, S., Ghorbel, F.: Robust and efficient Fourier-Mellin transform approximations for gray-level image reconstruction and complete invariant description. *Computer Vision and Image Understanding* 83, 57–78 (2001)
2. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*. Prentice Hall, Englewood Cliffs (2007)
3. Milanese, R., Cherbuliez, M.: A rotation-, translation-, and scale-invariant approach to content-based image retrieval. *J. Visual Comm. Image Rep.* 10, 186–196 (1999)
4. Osowski, S.: *Neural networks for information processing*. OWPW, Warsaw (2000) (in Polish)
5. Szczepaniak, P.S.: *Intelligent computations, fast transforms and classifiers*. EXIT Academic Publishing House, Warsaw (2004) (in Polish)
6. Stasiak, B., Yatsymirskyy, M.: Frequency Domain Methods for Content-Based Image Retrieval in Multimedia Databases. In: *Methods and Supporting Technologies for Data Analysis*, pp. 137–166. Springer, Heidelberg (2009)
7. Stasiak, B.: Fast Orthogonal Neural Network for Rotation-Translation- and Scale-Invariant Image Recognition. In: *Proc. of 12th IASTED International Conference on Intelligent Systems and Control* (2009)

Coarse-Grained Loop Parallelization for Image Processing and Communication Applications

Włodzimierz Bielecki¹ and Marek Palkowski²

Faculty of Computer Science, Technical University of Szczecin, 70210,
Żołnierska 49, Szczecin, Poland
e-mail: {wbielecki,mpalkowski}@wi.zut.edu.pl

Summary. Reducing time of application execution is significant for the quality of image processing and communication systems. Automatic coarse-grained parallelization of program loops is of a great importance for multi-core computing. This paper presents Iteration Space Slicing algorithms aimed at extracting coarse grained parallelism available in arbitrarily nested parameterized loops. We demonstrate that Iteration Space Slicing permits us to generate parallel code for image analysis, encoding and communication solutions. Experimental results are carried out with UTDSP benchmark.

1 Introduction

The increasing use of multi-core microprocessors necessitates the increasing need to expose coarse-grained parallelism available in sequential image processing and communication algorithms. The lack of automated tools permitting for exposing such parallelism decreases the productivity of parallel programmers and increases the time and cost of producing a parallel program.

Because most computations are contained in program loops, the automatic extraction of coarse-grained parallelism from loops is extremely important for multi-core systems, allowing us to produce parallel code from existing sequential applications and to create multiple threads that can be easily scheduled by a load balancers achieving a high system performance.

Loop parallelization is not trivial and dependence analysis is needed. Ignoring loop dependences causes that parallel code can produce not correct output. Two statement instances I and J are *dependent* if both access the same memory location and if at least one access is a write. I and J are called the *source* and *destination* of a dependence, respectively, provided that I is lexicographically smaller than J ($I \prec J$, i.e., I is always executed before J).

In this paper, we discuss Iteration Space Slicing algorithms extracting coarse-grained parallelism from program loops. We consider loops of image analysis, encoding, and communication application from the UTDSP

benchmark [1]. Experimental results are carried out in order to check the speedup and efficiency of generated parallel code.

2 Image Processing and Communication Algorithms in the UTDSP Benchmark

The UTDSP Benchmark Suite [2] was created in 1992 at the University of Toronto to evaluate the quality of code generated by a high-level language (such as C) compiler targeting a programmable digital signal processor (DSP). This evaluation was used to drive the development of specific compiler optimizations to improve the quality of the generated code and to modify the architecture of the target processor to simplify compiler's task. At that time, DSP applications were written in the assembly language of the target processor, making it necessary for us to write the applications ourselves in C [1].

A C compiler is used to translate target applications into machine operations that can be executed by the model architecture. The C compiler generates sequential code and performs register allocation based on the instruction set and the number of registers defined for the model architecture. A post-optimizer is then used to exploit the DSP-specific features of the model architecture. The post-optimizer also exploits parallelism in sequential code and creates executable code that runs on the model architecture.

Table 1. UTDSP kernel benchmarks

Kernels	Description
fft	Radix-2, in-place, decimation-in-time fast Fourier transform
fir	Finite impulse response (FIR) filter
iir	Infinite impulse response (IIR) filter
latnrm	Normalized lattice filter
lmsfir	Least-mean-squared (LMS) adaptive FIR filter
mult	Matrix Multiplication

The benchmark suite used in this study consists of six kernels and ten applications [1]. Table 1 shows kernel benchmarks, which consist of simple algorithms commonly used in DSP applications. The kernels usually constitute the inner loop of DSP applications; therefore the effectiveness of exploiting parallelism in kernels dominates the overall performance. In other words, the compiler must generate efficient code for kernels to maximize the utilization of the hardware resources in the model architecture. DSP kernels are simply small code fragments that represent important calculations in DSP applications. Examples of DSP kernels are: filters of various kinds, fast Fourier transform, matrix multiply. Table 2 shows the DSP application benchmarks,

Table 2. UTDSP application benchmarks

Applications	Description
G721_A, G721_B	Two implementations of the ITU G.721 ADPCM speech encoder
V32.modem	V.32 modem encoder/decoder
adpcm	Adaptive differential pulse-coded modulation speech encoder
compress	Image compression using discrete cosine transform (DCT)
edge_detect	Edge detection using 2D convolution and Sobel operators
histogram	Image enhancement using histogram equalization
lpc	Linear predictive coding speech encoder
spectral	Spectral analysis using periodogram averaging
trellis	Trellis decoder

which are commonly used in embedded systems. DSP applications, on the other hand, are entire programs that would be executed on a DSP in a commercial product. Using the suite of benchmarks with the application-driven design methodology thus makes the resulting model architecture an ideal design for embedded DSP processors. Examples of DSP applications are: linear predictive coding, image compression, an implementation of the G.721 modem protocol.

In this paper we apply Iteration Space Slicing to get parallel code for the UTDSP benchmarks.

3 Parallelism Extraction Using Iteration Space Slicing

Iteration Space Slicing (ISS) was introduced by Pugh and Rosser [3] as an extension of a program slicing proposed by Weiser [4]. It takes dependence information as input to find all statement instances that must be executed to produce the correct values for the specified array elements. A dependence graph refers to extensive set of dependence of a loop nest, described by dependence relations in the Presburger arithmetic. The algorithms presented in papers [6, 7, 8] show the usage of the Iteration Space Slicing for coarse-grained parallelization. Coarse-grained code is presented with synchronization-free slices or with slices requiring occasional synchronization. An (iteration-space) slice is defined as follows.

Definition 1. Given a dependence graph defined by set of dependence relations, a slice S is a weakly connected component of this graph, i.e., a maximal subgraph such that for each pair of vertices in the subgraph there exists a directed or undirected path.

Iteration Space Slicing (ISS) requires an exact representation of loop-carried dependences and consequently an exact dependence analysis which detects a dependence if and only if it actually exists. To describe and implement our algorithm, we chose the dependence analysis proposed by Pugh and Wonnacott [5] where dependences are represented by dependence relations.

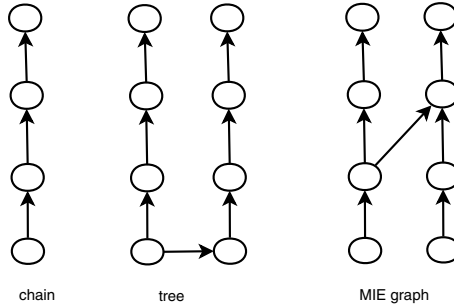


Fig. 1. Slices of the different topologies

A dependence relation is a tuple relation of the form $[input\ list] \rightarrow [output\ list]: formula$, where *input list* and *output list* are the lists of variables and/or expressions used to describe input and output tuples and *formula* describes the constraints imposed upon *input list* and *output list* and it is a Presburger formula built of constraints represented with algebraic expressions and using logical and existential operators.

We use standard operations on relations and sets, such as intersection (\cap), union (\cup), difference ($-$), domain ($\text{dom } R$), range ($\text{ran } R$), relation application ($S' = R(S): e' \in S' \text{ iff exists } e \text{ s.t. } e \rightarrow e' \in R, e \in S$), positive transitive closure of relation R , $R_+ = \{[e] \rightarrow [e'] : e \rightarrow e' \in R \vee \exists e'' , e \rightarrow e'' \in R_+ \vee e'' \rightarrow e' \in R_+\}$, transitive closure $R^* = R_+ \cup I$. In detail, the description of these operations is presented in [5, 10, 11].

Definition 2. An ultimate dependence source is a source that is not the destination of another dependence. Ultimate dependence sources and destinations represented by relation R can be found by means of the following calculations: $\text{domain}(R) - \text{range}(R)$.

Definition 3. The set of ultimate dependence sources of a slice forms the set of its sources.

Definition 4. The representative source of a slices its lexicographically minimal source.

We distinguish between the following topologies of slices:

1. *Single Incoming Edge (SIE)* graph, where each node has only one incoming edge. SIE graphs can represent the *chain* or *tree* topology.
2. *Multiple Incoming Edges (MIE)* graph, where one or more nodes has multiple incoming edges.

Our approach allows us to extract coarse grained parallelism represented with slices consists of the following steps:

1. find set S_{repr} of representative sources
2. reconstruct slices from their representatives and generate code scanning these slices.

The approach to extract synchronization-free slices [6] relies on the transitive closure of an affine dependence relation describing all dependences in a loop and consists of two steps. First, representatives of slices are found in such a manner that each slice is represented with its lexicographically minimal statement instance. Next, slices are reconstructed from their representatives and code scanning these slices is generated. Given a dependence relation R describing all dependences in a loop, we can find a set of statement instances, S_{UDS} , describing all ultimate dependence sources of slices as $S_{UDS} = \text{domain}(R) - \text{range}(R)$. In order to find elements of S_{UDS} that are representatives of slices, we build a relation, R_{USC} , that describes all pairs of the ultimate dependence sources that are transitively connected in a slice, as follows: $R_{USC} = \{[e] \rightarrow [e'] : e, e' \in S_{UDS}, e \neq e', (R^*(e) \cap R^*(e'))\}$.

The condition ($e \neq e'$) in the constraints of relation R_{USC} means that e is lexicographically smaller than e' . Such a condition guarantees that the lexicographically smallest element from e and e' will always appear in the input tuple, i.e., the lexicographically smallest source of a slice (its representative source) can never appear in the output tuple. The intersection ($R^*(e) \cap R^*(e')$) in the constraints of R_{USC} guarantees that elements e and e' are transitively connected, i.e., they are the sources of the same slice.

In second step we can use the following algorithms in order to reconstruct slices:

- **Gen_affine** [6] - an algorithm allowing us to generate code changing synchronizationfree slices of an arbitrary topology if R^* has a closed form and affine constraints. This algorithm uses well - known code generation techniques to scan elements of affine sets representing synchronization-free slices. It requires computation of R^* .
- **Gen_graph** [7] an algorithms allowing us to scan slices of the chain, tree or graph topology that does not require computation of R^* . Within each slice, it employs free-scheduling of statement instance (statement instances are executed as soon as their operands are available). It can be applied to parallelize loops for which Gen_affine fails to extract synchronization-free slices because of impossibility to compute affine R^* or as an alternative for Gen_affine aiming at increasing code locality.
- **Synch_MP** [8] - an algorithm employing synchronization based on message passing using both OpenMP and POSIX locks functions [9]. When above algorithms extract only a single slice, this algorithm can be used to seek for slices requiring synchronization. The generated code consists of send and receive functions to synchronize the slices execution.

4 Experiments

To extract dependences, the Petit tool [11] was used. Petit was able to find dependences in 34 UTDSP loops. For 34 UTDSP loops qualified to experiments, Iteration Space Slicing allows us to extract slices for 20 UTDSP loops. Table 3 shows the numbers of parallelized loops for UTDSP kernels and applications.

To check the performance of parallel code, speedup and efficiency was examined. Speed-up is a ratio of sequential time and parallel time, $S=T(1)/T(P)$, where P is a number of processors. Efficiency, $E=S/P$, tells us about usage of available processors by parallel code.

Three loops from the Benchmark were examined, two loops from *Edge_detect* application and one from *Compress* application. For each loop multiple sources and transitive closure of all relations were found. It allows

Table 3. Parallelized loops of UTDSP kernels and applications

Application	Gen_affine	Gen_graph	Gen_affine
compress	3	3	1
edge_detect	3	2	1
fir	1	1	0
g721	1	1	0
fir	1	1	0
lpc	6	6	0
histogram	1	1	0
spectral	1	1	0
mult	2	2	0
trellis	1	1	0
summary	20	19	2

Table 4. Speed-up and efficiency of loops from the UTDSP benchmarks

Loop:	Params	2 CPU		4 CPU		8 CPU	
		S	E	S	E	S	E
Edge_detect_1	N = 1000	1,19	0,59	1,48	0,37	1,95	0,24
	N = 2000	1,47	0,74	2,25	0,56	3,42	0,43
	N = 3000	1,54	0,77	2,70	0,68	3,84	0,48
Loop:	Params	2 CPU		4 CPU		8 CPU	
		S	E	S	E	S	E
Edge_detect_4	N = 500, K=10	1,64	0,82	3,02	0,76	4,58	0,57
	N = 1000, K=10	1,72	0,86	3,35	0,84	5,99	0,75
	N = 1500, K=10	1,73	0,87	3,43	0,86	6,57	0,82
Loop:	Params	2 CPU		4 CPU		8 CPU	
		S	E	S	E	S	E
Compress_1	N = 1024	1,43	0,72	1,82	0,45	3,04	0,38
	N = 1536	1,77	0,88	2,56	0,64	3,58	0,45
	N = 2048	1,78	0,89	3,30	0,82	4,82	0,60

us to use Gen_affine algorithm to generate parallel code for all loops. Table 4 shows speed-up and efficiency for 2, 4, and 8 processors. Experiments were carried with Workstation Board S5000XVN Intel Xeon Quad Core, 1.6 Ghz, 8 CPU (2 quad core CPU with cache 4 MB), 2 GB RAM, Ubuntu Linux.

5 Conclusion

In this paper, we demonstrated that ISS extracts coarse-grained parallelism for image processing and communication applications. Loops of the UTDSP benchmarks are divided on many slices which are mapped to processors as threads. Coarse-grained parallelism advantage is no synchronization or occasional synchronization between threads. It allows us to achieve significant speed-up of parallel programs on popular memory-shared machines with multi-core processors.

ISS permits us to parallelize loops for which other popular techniques, like Affine Transformation Framework or Unimodular Transformations, fail to produce parallel code [6].

In the future, we are going to examine other benchmarks of image analysis, video encoding and communication using the Iteration Space Slicing framework. We also plan to improve dependence analysis for loops written in C and design automatic algorithms improving locality of generated code for a particular parallel machine.

References

1. Peng, S.H.: UTDSP: A VLIW Programmable DSP Processor. Department of Electrical and Computer Engineering, Thesis, University of Toronto (1999)
2. UTDSP benchmark suite, <http://www.eecg.toronto.edu/~corinna/DSP/infrastructure/UTDSP.html>
3. Pugh, W., Rosser, E.: Iteration space slicing and its application to communication optimization. In: International Conference on Supercomputing, pp. 221–228 (1997)
4. Weiser, M.: Program slicing. IEEE Transactions on Software Engineering, 352–357 (1984)
5. Pugh, W., Wonnacott, D.: An exact method for analysis of value-based array data dependences. In: The Sixth Annual Workshop on Programming Languages and Compilers for Parallel Computing, Springer, Heidelberg (1993)
6. Beletska, A., Bielecki, W., San Pietro, P.: Extracting coarse-grained parallelism in program loops with the slicing framework. In: ISPD 2007 Proceedings of the Sixth International Symposium on Parallel and Distributed Computing, vol. 29, IEEE Computer Society, Washington (2007)
7. Bielecki, W., Beletska, A., Palkowski, M., San Pietro, P.: Extracting synchronization-free trees composed of non-uniform loop operations. In: Bourgeois, A.G., Zheng, S.Q. (eds.) ICA3PP 2008. LNCS, vol. 5022, pp. 185–195. Springer, Heidelberg (2008)

8. Bielecki, W., Palkowski, M.: Using message passing for developing coarse-grained applications in OpenMP. In: Proceedings of International Conference on Software and Data Technologies ICSoft 2008 (2008)
9. OpenMP API, <http://www.openmp.org>
10. Kelly, W., Maslov, V., Pugh, W., Rosser, E., Shpeisman, T., Wonnacott, D.: The omega library interface guide. Technical report, College Park, MD, USA (1995)
11. Pugh, W., Wonnacott, D.: An exact method for analysis of value-based array data dependences. In: The Sixth Annual Workshop on Programming Languages and Compilers for Parallel Computing, Springer, Heidelberg (1993)

SMAS - Stereovision Mobility Aid System for People with a Vision Impairment

Rafał Kozik

Institute of Telecommunications, University of Technology & Life Sciences,
Kaliskiego 7, 85-796 Bydgoszcz, Poland
e-mail: rafal.kozik@upt.edu.pl

Summary. New computer vision solutions dedicated for blind and partially sighted people have been recently introduced as a result of significant progress in computer science. Also the growing computation power of mobile and portable devices together with development of information systems allow to adopt and apply new and robust solutions that are able to work in nearly in a real-time and share and use information spread over IP network. Many of currently developed solutions are dedicated to support the user, giving the information about divert obstacles located in the environment. However many of them are using simple detectors (commonly ultrasonic echo-location) for obstacles tracking without its classification and recognition. Therefore the solution presented in this paper engages the stereo camera and image processing algorithms to facilitate its user with object detection and recognition mechanisms. The inference engine combined together with ontology based problem modeling allows to handle the risk, predict possible user's moves and provide the user with appropriate set of tips that will eliminate or reduce the discovered risk.

1 Introduction and Motivation

Current solutions used by people with a vision impairment engages simple object detection techniques such as ultrasonic echo-location and are often recommended to be used as secondary aid tool which is not intended to replace the traditional aids such as white cane and guide dog. Some of already used commercial solutions (e.g "Miniguide ultrasonic mobility aid" [2]) are able to identify a small set of objects. However the common drawback of such solutions is a lack of possibility of enlarging the amount of recognized instances. Commonly these allows to avoid obstacles, locate doorways, gaps or determine if elevator doors are open. However in the last few years the research concerning the alternatives solutions supporting blind people has been growing and introducing new designs and systems that focus on adaptive and intelligent methods utilizing stereo cues and image processing approaches [1].

2 System Architecture

The general architecture of the proposed system is presented in Fig. 1. It uses two standard cameras as a stereovision system. These are bind together and calibrated prior to the use. These allows to obtain the depth map, which is essential to get the information about the position and size of nearby localized obstacles thanks to depth map segmentation algorithm. Further processing allows to extract distance between particular obstacle and camera. For each detected obstacle the feature vector is built and is used to query the database to identify that object. When object is identified the Ontology Based Recognition System (OBRS) is queried to identify further object properties and interconnected threats. The semantic description allows to:

- classify particular obstacle to particular class of object
- provide description of relations between objects
- provide description of relations between particular object and the environment it is located.
- model dynamic relations between objects by providing the semantic rules
- create risk management

The OBRS is used both for risk assessment and proposing appropriate action plan, which will reduce or eliminated discovered threat. Finally the message describing the threats and suggested action plan is build in textual form and converted to voice via TTS (Text To Speech) application.



Fig. 1. Generalized system architecture

3 Obstacles Detection

The images obtained from stereocamera allow to reconstruct the 3D structure of observed scene. However to do this accurately the stereo cues have to be calibrated to obtain the intrinsic camera parameters and produce transformation matrices that will rectify images coming from the left and the right cameras. Also knowing the intrinsic properties (such as focal length) it is

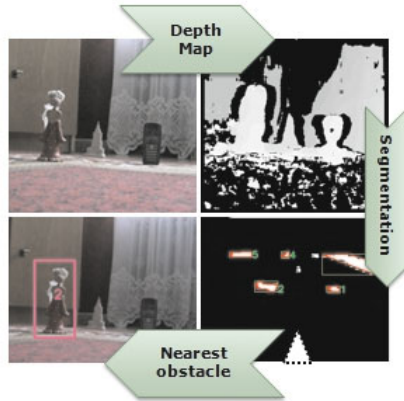


Fig. 2. Depth map segmentation and obstacles extraction

possible to estimate the distance to particular obstacle via triangulation. Applying this algorithm to all pixels in left image and its correspondences in right image the depth map is obtained. However the most challenging thing in the whole procedure is to find the corresponding pixels in both images. To extract the obstacles located in front of a user the depth map segmentation algorithm is used. The procedure is shown in Fig. 2.

3.1 Stereo Matching

There are different approaches to stereo matching. These can be generally divided into two groups. First group of algorithms use the local features to compute the disparity map (the product of stereo matching). The second group engages both local and global constraints, minimizing the cost function. Both of these methods have advantages and disadvantages. The local-based method allows to obtain the disparity map in relatively short time, but its quality (accuracy) is quite poor. In the other hand the methods engaging global constraints allows to generate high accuracy disparity maps, however not in a real time. The accuracy of disparity map estimation of selected stereo matching algorithms, for example images, are shown in Fig. 3.

3.2 Depth Map Segmentation

Before the particular object can be located in the environment it has to be segmented and extracted from the depth map. However this is not trivial task to do since different object may occupy different depth ranges, which makes segmentation more difficult. The main idea which solves this problem aims at generating the bird's eye view image of the scene as is it shown in Fig. 2. This can be obtained by applying the formula 1 to each pixel of the image,

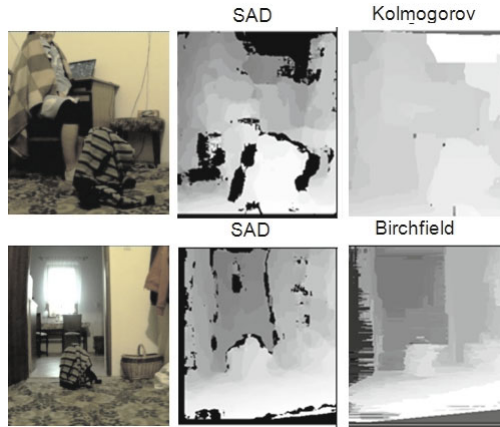


Fig. 3. Examples different algorithms used for stereo matching (SAD (sum of absolute distances), Kolmogorov [4], Birchfield [5])

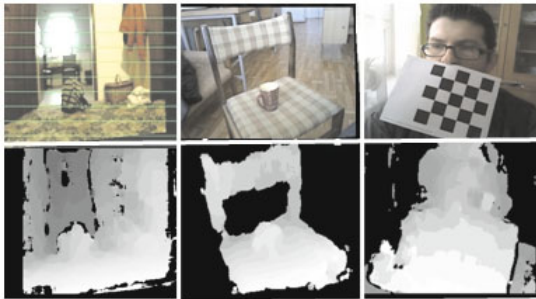


Fig. 4. Examples of depth map obtained via SAD algorithm

where *height* indicates the depth map height, and $depthMapPixEq(x, j, y)$ is equal 1, if pixel at position (x, j) has depth y and 0 otherwise.

$$birdsEyeView(x, y) = \sum_{j=0}^{height} depthMapPixEq(x, j, y) \quad (1)$$

The bird's eye view image is further segmented and each of segments is labeled to identify the width and the depth ranges of the object (Fig. 5C). Having this information the obstacle can be easily located in the original image as it is shown in Fig. 5D. This region of texture is used as an input to obstacle identification. Examples of algorithm at work are presented in Fig.6.

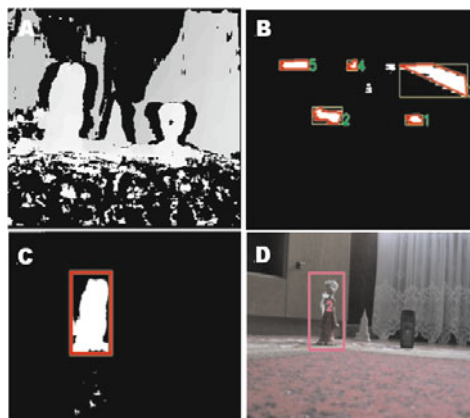


Fig. 5. Obstacle segmentation procedure. A - depth map, B - Thresholded Bird's Eye View, C - Filtered depth map and segmented obstacle, D - localized obstacle in original image

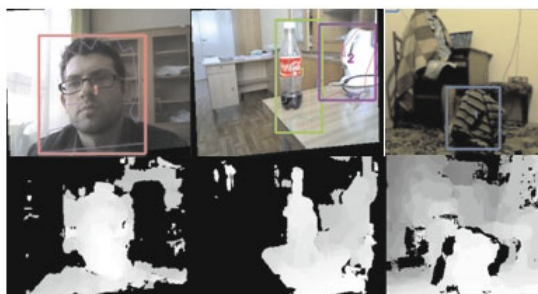


Fig. 6. Example of segmented objects

3.3 Object Identification

There are different approaches to index images in large data base. However the most promising and yielding high effectiveness and robustness to rotation and occlusions, are methods named BOF (Bag Of Features). These extract characteristic points (typically maxima and minima of luminance) from the ROI (Region Of Interest). Each of such points is described by individual feature vector. In this case the SURF descriptors were used. Using this set of features (points with descriptors) the database of known object is queried to find the closest match, which matching cost is lower than system threshold. The brute data base search (computing distances to all instances in data base) is quite computationally expensive, therefore using more efficient data structure such as inverted file system allows to obtain the query results significantly faster. Also (as showed Sivic and Zisserman in [6]) adopting the method of visual words increases searching efficiency. According to

this approach the local descriptor are quantized into visual word (commonly k-means clustering algorithm is adopted for that task). For particular image a histogram of visual words is created. This allows to compare histograms (rather than descriptors) to find closest matches.

3.4 Risk Management

The presence of obstacle in front of the user does not implies the awareness of all threats connected with that object. In real worlds there are complicated relations between object causing the threats to show up only if two or more objects appear together. Particular the mug of hot coffee on top of chair would be classified as a single obstacle and as a result the risk could be underestimated by blind person. Therefore, proposed in this paper method of extracting texture information, would simply recognize that there are two objects. Furthermore different obstacles of the same sizes may expose blind person to different threats. However simple object identification is also insufficient (e.g. knife should not be considered as threat if blind person operates it and is aware of that fact). Therefore to handle these complicated relations between object and to provide good semantic description the ontology notation model is use. The simplified model showing high level relations and the main concepts is shown in Fig.7

The ontology focuses on risk assessment. At the beginning there is an recognized object, which has its properties. In example mentioned above coffee mug will have property "may contain hot coffee". With that property there are connected some threats. In this example there are two of them: "burn yourself" and "spill content". Each of theses threats increases the whole risk the blind person is exposed to. However for each of treats there may be some reaction (focused on the object) that will eliminate or reduce the risk.

3.5 Experiments

There are two scenarios of experiments. The first scenario aims at evaluating the functionalities responsible for risk assessment and informing the blind

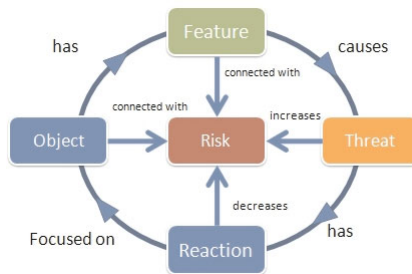


Fig. 7. The simplified ontology diagram




	Scenario	System response:
	Threats identification	Object: Coffee Mug Features: May contain hot coffee Threats: Burn yourself, spill content Reaction: Be careful
	Object searching and recognition	Object: Phone
		Object: Owner

Fig. 8. Examples of experiments

person about threats and appropriate reaction minimizing the risk. The second scenario aims at evaluating object identification and object searching functionalities. Examples of results (including system reaction in textual form) are presented in Fig. 8.

3.6 Conclusions

In this paper prototype of mobility aid system for people with vision impairment was described. The tentative results are promising and show that engaging the stereovision systems together with robust object extraction algorithm and ontological description of the environment allow to assess the risk the blind person is exposed to. Furthermore proposed here distributed architecture allows to perform huge amount of operation in nearly real time. What is more such architecture gives opportunity to seamless system re-learning (by adding new instances to distributed databases) and reconfiguration without maintaining the user's personal computer.

The further work focuses on adopting more robust stereo matching algorithms that will yield satisfactory depth map. This will allow extract the obstacles more accurately. Also the object identification functionalities have to be supported by other recognition techniques. The BOF-based methods are promising and fast, however there are many objects in the environment which are hard to identify using the texture. However these usually have distinctive color and shape. What is more the BOF allows to recognize the particular instance but not class of object. In example there are plenty of chair instances that we haven't seen before but are still capable to classify it as chairs, according to their common physical features. Therefore decomposing the object into low level features and using the ontology to describe the relations between them is expected to cover that problem.

References

1. Pelcztnski, P.: Travel Aid System for the Blind. *Image Processing and Communications Challenges*, 324–333 (2009)
2. The Miniguide project homepage, <http://www.gdp-research.com.au>
3. Sun, J., Li, Y., Kang, S., Shum, H.: Symmetric stereo matching for occlusion handling. In: *CVPR*, pp. II 399–406 (2005)
4. Kolmogorov, V., Zabih, R.: Computing visual correspondence with occlusions via graph cuts. In: *I*, pp. II 508–515 (2001)
5. Birchfield, S., Tomasi, C.: Depth Discontinuities by Pixel-to-Pixel Stereo. *International Journal of Computer Vision* 35(3), 269–293 (1999)
6. Sivic, J., Zisserman, A.: Efficient Visual Search for Objects in Videos. *Proceedings of the IEEE* (2008)
7. Kirasic, D., Basch, D.: *Ontology-Based Design Pattern Recognition*. LNCS (2009)
8. Francois, A.R.J., Nevatia, R., Hobbs, J., Bolles, R.C.: VERL: an ontology framework for representing and annotating video events. *IEEE MultiMedia* 12(4), 76–86 (2005)
9. Latfi, F., Lefebvre, B., Descheneaux, C.: Ontology-based management of the telehealth smart home, dedicated to elderly in loss of cognitive autonomy. In: *Proceedings CEUR Workshop*, vol. 258 (2007)
10. Fergus Torralba, R., Weiss, Y.: Small codes and large databases for recognition. In: *CVPR* (2008)
11. Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Lost in quantization: Improving particular object retrieval in large scale image databases. In: *CVPR* (2008)

Extracting Symbolic Function Expressions by Means of Neural Networks

Jarosław Majewski and Ryszard Wojtyna

Faculty of Telecommunication and Electrical Engineering

University of Technology and Life Sciences Bydgoszcz, Poland

e-mail: Jaroslaw.Majewski@utp.edu.pl, Ryszard.Wojtyna@utp.edu.pl

Summary. In this paper, a new neural network capable of extracting knowledge from empirical data [1]–[6] is presented. The network utilizes the idea proposed in [2] and developed in [3, 4]. Two variants of the network are shown that differ in relationships describing activation functions of neurons in the network. One variant utilizes logarithmic and exponential functions as the activation ones and the other is based on reciprocal activation functions. The first network variant is similar that proposed in [3]. The difference is that in our network the logarithmic activation function works with hidden layer neurons while in [3] with input signals. In the second variant, all activation functions are of $1/x$ type. To the author’s knowledge, such a network has not been published in the literature so far. Like that of [3], our network provides a real valued symbolic relationship between input and output signals, resulting from numerical data describing the signals. The relationship is a continuous function created on the basis of a given set of input–output numerical data when learning the network. Extraction of the symbolic function expression is carried out after the training is finished. By forming the symbolic expression, the neural network structure and synaptic connection weights associated with the neurons are taken into account. The ability of knowledge extraction, also called law discovery, is a consequence of applying proper activation functions of neurons included in hidden and output layers of the network. The neural network under consideration can also play an inverse role to the above mentioned. Instead of extracting the symbolic relation, it can also be used as a neural realization of continuous functions expressed in a symbolic way. The presented theory is illustrated by an example.

1 Introduction

Determining continuous–function relationships between two empirical data sets can be realized using neural networks [1]–[6], where one of the set is regarded as the network input data and the other as its output one. In case of applying neural networks, solving this task means automatic knowledge extraction using the networks. The neural network role is to discover laws

and regularities governing the input–output data relations. Moreover, the network should enable to extract and write the laws in a symbolic form.

There are two approaches to the problem under consideration. The first one consists in reading decision rules of an IF ... THEN ... type from the neural network trained by means of the empirical input–output data set, and next writing the rules. This is, in fact, a neural realization of expert systems. In the other approach, network Boolean functions are extracted from the trained network and written. In both cases, either sigmoid or certainty–factor–based (CF–based) relationships are used as activation functions [1, 6]. Moreover, for both approaches one applies an adaptive way of transforming a knowledge included in the input–output data into weights between neurons (network training). To extract the knowledge and write it in symbolic forms after finishing the adaptation learning, special techniques are applied [1, 6].

There are two ways to solve the knowledge extraction task. In the first method, called open–box approach or sometimes decomposition approach, each internal element of the network is examined. Making use of full information about the network structure and its neuron weight values, one obtains the searched logic rules. The second method is based on using only an observed input–output behavior of the network to determine the rules. Such an approach is called black–box or pedagogical approach [5].

The network proposed in this paper belongs to the second group of extraction techniques and presents an improved approach to the problem. An advantage of our approach is that it enables to extract real–valued continuous functions describing relationships between input and output system data. The continuous function extraction is realized in the learning adaptive process of the network. It is worth stressing that information about the continuous functions can be read out, after the learning is finished, from the network topology and values of the neuron synaptic connections (neuron weights) and can be written in a symbolic form using the open–box method.

The first network proposed in this paper resembles the nets presented in [3]. The similarity is that in both networks natural–logarithm and exponential functions are used as activation functions. The difference between our approach and that of [3] is that the activation functions concern different network layers. In case of [3], finding natural logarithm concerns the network input data vector \mathbf{x} , and the exponential operation is performed on its hidden–layer output signals. This is illustrated in Fig. 1. Our network is shown in Fig. 2. In the proposed solution, finding natural logarithms is performed on output signals of the network hidden layer, while the exponential operation concerns the network output neuron (only one).

In Fig. 1, a neural network is shown which corresponds to the relation proposed in [3] and is described by:

$$y_t = c_0 + \sum_{i=1}^h c_i \exp \left(\sum_{j=1}^n w_{ij} \ln(x_{tj}) \right), \quad (1)$$

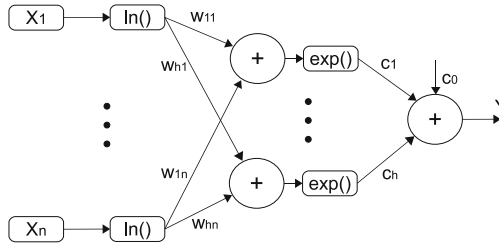


Fig. 1. Topology of neural network proposed in [3], working according to the expression given by (1)

where y_t is a target value at the network output, x_{tj} are components of the input vector corresponding to y_t , n is the number of input neurons, w_{ij} are weights between input neurons and hidden ones, h is the number of hidden neurons, c_i are weights between the hidden neurons and output node and c_0 is a real valued coefficient.

From (1) and from Fig. 1 it is seen, that the number of variables, excluding c_0 , whose values are changed in the learning process equals a product of n and h , where n is a given number, while h an unknown number.

In the next section, a new simple network is presented that is also based on logarithmic and exponential activation functions but the activation functions operate in different layers compared to the network shown in Fig. 1. Another solution of the knowledge extracting task, to be proposed by us, is network shown in section 3. It is based on reciprocal activation functions, i.e. functions of a $1/x$ type. An example illustrating usefulness of our proposals is given in section 4, and concluding remarks in section 5.

2 Simple Network Using Logarithmic and Exponential Functions

Basic diagram of the proposed neural network is show in Fig. 2. It consists of input layer (N input nodes denoted by X_i), one hidden layer with M neurons (any number of them) and one output neuron.

Call all hidden layer neurons in Fig. 2 by *logotron* because their activation function is given by natural logarithm function, $g(x) = \ln(x)$, and the output neuron by *expotron* as its activation function is expressed by exponential function $f(x) = \exp(x)$. Unlike a classical perceptron, which is able to extract knowledge from input–output data in a Boolean function form only, our network provides the extracted knowledge in a real–valued continuous function way. The reason for introducing the logotron and expotron names is to emphasis that such activation functions are not used in typical neural networks and to simplify description of more complex networks built of networks of the type shown in Fig. 2. As a result, one can extract a more complex function expressions than that discussed in this paper.

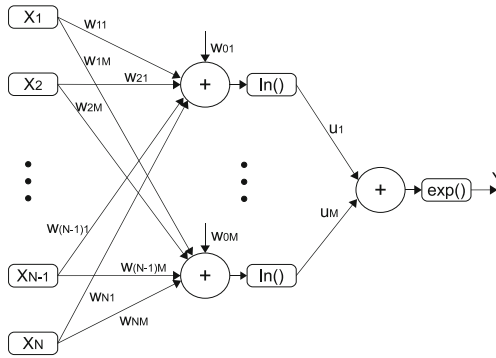


Fig. 2. Proposed basic neural network, utilizing exponential and logarithmic functions, suitable for extracting symbolic knowledge about input–output relationships

Consider properties of the network shown in Fig. 2. For N inputs and M hidden–layer neurons, output signal is given by:

$$y = \exp \left(\sum_{m=1}^M u_m \ln \left(w_{0m} + \sum_{n=1}^N w_{nm} x_n \right) \right), \quad (2)$$

where u_m represents weights between hidden neurons and output neuron, and w_{nm} weights between input nodes and hidden layer neurons.

Equivalently, the relationship (2) can be expressed as:

$$y = \prod_{m=1}^M \left(w_{0m} + \sum_{n=1}^N w_{nm} x_n \right)^{u_m} \quad (3)$$

As can be seen from (3), our network realizes in a direct and precise way both weighted adding (second component in the bracket) and multiplication of the brackets raised to u_m power.

Taking into account the network ability to learn in an adaptive way, i.e. vary the w_{nm} and u_m weights between neurons, as well as to generate the acquired knowledge, it is possible to train it in such a way to realize any function of the form given by (2). This is true provided that a finite set of empiric pattern points (\mathbf{x}, y) , satisfying the realized function, is used in the training process. The number of the set elements, necessary for obtaining satisfactory training results, can be different, depending on the case considered, but usually is not very large. It is possible to carry out a successful learning making use of a few pattern points only.

An important advantage of the network of Fig. 2 is that it is trained using a finite set of pattern points given by numerical data. Moreover, after the training is finished, the obtained final function is a symbolic relationship. This makes it possible to extract knowledge in continuous function forms on the basis of input–output data whose properties and laws are unknown.

3 Network Based on Reciprocal Activation Functions

The network presented in Fig. 2 includes logarithmic and exponential activation functions and realizes multiplication and rising-to-power operations, according to (3). Consider the network shown in Fig. 3 with all activation functions of a reciprocal type. Like that of Fig. 2, it is a three layer network with only one output neuron. This network realizes another interesting function, namely the relationship given by:

$$y = \frac{1}{\left(\sum_{n=1}^N \frac{u_n}{x_n}\right)} \tag{4}$$

The relation (4) can be utilized, for example, to describe in simple manner equivalent impedance representing K impedances connected in parallel. This is achieved when K weights, u_n , equal one, and the others are equal to zero. Regarding the network output signal as the equivalent impedance, $y = Z$, and input signals as the impedances connected in parallel, $x_n = Z_n$, we obtain:

$$Z = \frac{1}{\left(\sum_{i=1}^K \frac{1}{Z_i}\right)} = \left(\sum_{i=1}^K \frac{1}{Z_i}\right)^{-1} \tag{5}$$

The expression (5) takes a simple polynomial form when $K = 2$:

$$Z = \frac{Z_1 Z_2}{Z_1 + Z_2}, \tag{6}$$

and becomes more and more complex when K increases. For example, if $K = 3$ we get:

$$Z = \frac{Z_1 Z_2 Z_3}{Z_1 Z_2 + Z_1 Z_3 + Z_2 Z_3} \tag{7}$$

Acting according to the idea described in [2], a neural realization of the expression (7) requires using one network described by (1) with three inputs,

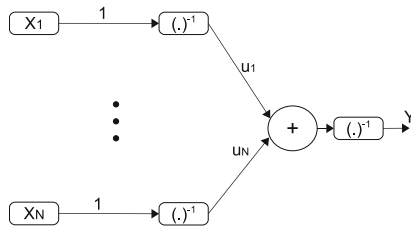


Fig. 3. Basic neural network (parallel-like unit) with reciprocal activation functions in hidden and output layers

one so called *product unit* defined in [2] and, additionally, one reciprocal-function operation. An advantage of our network is that the same task can be realized much easier, i.e. using one network of the type shown in Fig. 3 with 3 inputs and the same number of components as that included in the triple-input product unit proposed in [2].

In (4), u_n can be regarded as weight of a given component in the denominator sum. From mathematical point of view, the lower is u_n , the lower is the component influence on the network final result. In case of binary weights, a given input signal, x_k , is processed (for $u_k = 1$) or not (for $u_k = 0$). The network learning consists in modifying the u_n value. The knowledge extraction (extraction of final expression) means reading the u_n values out and formulating, on this basis, symbolic relationship describing the network operation.

We propose to call the network of Fig. 3 *parallel-like unit*. By analogy to impedances connected in parallel, we also propose to present the relationship (4) using the below given symbolic form:

$$y = \frac{x_1}{u_1} \parallel \frac{x_2}{u_2} \parallel \dots \parallel \frac{x_N}{u_N}, \quad (8)$$

which significantly simplifies symbolic expressions characterizing more complex networks with parallel-like units.

As for learning the parallel-like units, the training process is like in typical neurons. This enables connecting them with neurons of other types, for instance with product units and linear units. Such a connection leads to more complex networks of greater possibilities. In other words, it is possible to extract more complex symbolic expressions. Networks of this type are simpler and can be trained faster compared to other ones.

4 Learning the Network

Mathematically, the expression (4) describing the network shown in Fig. 3 which we have called parallel-like unit, is correct. The question, however, is whether there exists a convergent algorithm to train such a network. To answer this question, a training process using a standard Back Propagation (BP) method has been applied. Vectors used to train the network are shown in Table I. In this table, we have twelve vectors with 4 components of each vector (x_1, x_2, x_3, y) .

The vectors given in Table I satisfy the expression:

$$y = \frac{x_1}{2} \parallel \frac{x_2}{3} \parallel \frac{x_3}{3.5} \quad (9)$$

Comparing (8) with (9) it is seen that perfect training should lead to $u_1 = 2, u_2 = 3$ and $u_3 = 3.5$. The performed training turned out to be a convergent process. After carrying out 257413 epochs, for learning coefficient equal to 0.0006 and momentum coefficient equal to 0.2, we obtained

Table 1. 4-dimensional vectors used to train the network of Fig. 3.

x_1	x_2	x_3	y
0.100	0.500	1.000	1.200
1.400	2.000	2.300	3.000
4.100	5.000	3.600	7.700
1.000	4.400	2.345	0.230
0.345	4.700	9.200	3.010
3.200	0.230	0.100	4.340
23.000	12.300	4.500	3.400
-9.200	-2.300	-4.500	-23.000
-0.300	2.560	1.453	7.320
0.040	0.201	0.246	0.064
0.103	8.579	2.393	0.662
-0.098	0.068	0.030	0.700

$\hat{u}_1 = 2.0, \hat{u}_2 = 2.997$ and $\hat{u}_3 = 3.498$. Further learning resulted in the \hat{u}_1, \hat{u}_2 and \hat{u}_3 values being even closer to the expected values 2, 3 and 3.5.

Having in mind that to train the product unit presented in [3, 4] not standard but a special algorithms like Random Search, Particle Swarm Optimization, Genetic Algorithms, Leapfrog Optimization and BFGS-like algorithm called BPQ were applied, we expect to obtain still better results of learning the network of Fig. 3 when applying a specialized algorithm suitable for this task.

5 Conclusions

A problem of extracting knowledge from empirical data by means of neural networks is considered in the paper. The received knowledge has a symbolic relation form and results from training triple-layer network which includes proper activation functions. Like in typical neuron networks, the training process provides values of weights between the network neurons. The weights are parameters of the resulting continuous function expressions relating two empirical data sets. One of the sets is used as input data of the trained network and the other as the network output data.

Two network structures have been proposed that are attractive to realize the knowledge discovery task. One is based on using several natural-logarithm activation functions in hidden layer and an exponential activation function associated with the network single-output neuron. Its scheme is shown in Fig. 2 and symbolic function describing the input-output relation is given by (2). The structure of Fig. 2 is similar to that proposed in [3] and shown in Fig. 1, where several natural-logarithm activation functions are used in the network input layer and several exponential activation function in its hidden layer. Like the network of Fig. 1, our proposal shown in Fig. 2 can equivalently be expressed as a function of products of the input signals, which is given

by (3). This makes it an attractive basic unit useful for creating complex relations in a polynomial form.

Apart from the network with the logarithmic and exponential functions, a new network based on reciprocal activation functions, placed in hidden and output layers, has been proposed as well. Its structure is shown in Fig. 3, while basic relationship given by (4) and, equivalently, by (8). We have called it parallel-like unit because it provides a continuous function relationship suitable, for example, to describe, in a simple and compact form, equivalent impedance representing several impedances connected in parallel. As in the previous network of Fig. 2, parameters of the function (4) result from proper training the network. One training example has been presented which shows that a convergent learning can be easily performed.

Combining the networks of Fig. 2 with that of Fig. 3, a lot of interesting continuous function relationships, whose parameters are attainable by learning the networks, can be determined. This is a direction of our further research in this field.

References

1. Fu, L.M.: Knowledge Discovery by Inductive Neural Networks. *IEEE Trans. On Knowledge and Data Engineering* 11(6) (November/December 1999)
2. Durbin, R., Rumelhart, D.: Product Units: A Computationally Powerful and Biologically Plausible Extension to Backpropagation Networks. *Neural Computation* 1, 133–142 (1989)
3. Saito, K., Nakano, R.: Law Discovery using neural networks. In: *Proc. of the 15th International Joint Conference on Artificial Intelligence*, pp. 1078–1083 (1997)
4. Ismail, A., Engelbrecht, A.P.: Paining Product Units in Feedforward Neural Networks using Particle Swarm Optimization. In: Bajic, V.B., Sha, D. (eds.) *Proceedings of the International Conference on Artificial Intelligence Development and Practice of Artificial Intelligence Techniques*, Durban, South Africa, pp. 36–40 (1999)
5. Tickle, A.B., Andrews, R., Golea, M., Diederich, J.: The Truth Will Come to Light: Directions and Challenges in Extracting the Knowledge Embedded Within Trained Artificial Neural Networks. *IEEE Trans. on Neural Networks* 9(6) (November 1998)
6. Fu, L.M.: Learning in Certainty-Factor-Based Multilayer Neural Networks for Classification. *IEEE Trans. on Neural Networks* 9(1) (January 1998)

Mathematical Morphology in the Process of Musical Notation Recognition

Arkadiusz Rajś

Telecommunication Institute, University of Technology and Life Sciences,
Kaliskiego 7, 85-791 Bydgoszcz
e-mail: arajs@utp.edu.pl

Summary. Mathematical Morphology is a tool for extracting image components that are useful for representation and description. This article shows part of process of automatic Optical Music Recognition.. It suggest effective methods used to remove staff line and preparing image to symbol identification. This method based on mathematical morphology. The experimental results are showing.

1 Introduction

The automatic Visual systems have been developed and upgraded for many years to increase the domain of their use. At the same time the tendency of storing the information as a digital one leads to realization of such systems of storing the pictures which let not only change the pictures from analog to digital system but also allow to correct its quality.(for example hum reduction in the image). The optical music recognition systems are the example of such ones. They are used in digitalization of music notation and in changing the handwritten music notes into print and also in putting notes into devices reproducing sounds automatically.

The automatic systems of recognizing musical notation are constructed basing on four functional modules which realize:

- detection of staff position and its removal
- detection of position of musical symbols basing on the identified staff from the previous point
- identification of the symbols
- assignation of the semantic relation between the music symbols and its notation accepted by existing music editors.

Mathematical morphology is the basic part of digital image processing . It seems worthy to analyze possibility of using the operations to build the systems recognizing the music notation.

The aim of this article is to describe the ways of using erosion and dilatation to detect musical symbols

2 Staff Line Removal Method

The aim of the process of removal staff line from musical notation is to prepare the image to detect the musical symbols. In the automatic recognition of the musical notation the staff line is the element which might aggravate the correct position of the symbols (lines might be read as elements of musical notation). In later process of their identification the lack of the staff does not make any problem because its position can be memorized or the position of different symbols can be compared with the g or f clef (placed at the beginning of the staff line).

The form of the digital image containing the musical notation should have the following characteristics:

- as the result of the canvassing of the image the position of the staff line must be horizontal because any alteration in angle in its position makes the removal of the staff line its further placing and the identification of the symbols impossible
- bit depth of the image containing musical notation should one bit (set of black and white pixels)
- it should be assumed that the changes will be done in black pixels and the white ones will create the background.

The task of removing the staff line from the image can be done in two stages - realizing the erosion operation and then dilatation one.

Mathematically the erosion operation for the monochromatic symbols can be written as following [2, 3]:

$$E_{mono}(L, SE) = \min_{m,n \in SE} L(m, n) = \min_{SE}(L) \quad (1)$$

where: $L(m, n)$ - brightness of the point about coordinate (m, n) , SE - structural template for erosion.

By analogy mathematically written operation of dilatation for monochrome objects is expressed as [2]:

$$D_{mono}(L, SE) = \min_{m,n \in SE} L(m, n) = \min_{SE}(L) \quad (2)$$

where: $L(m, n)$ - brightness of the point about coordinate (m, n) , SE - structural template for dilatation.

The compound of the erosion operation and then dilatation one is called the opening operation however in this case the structural elements SE are not identical for the operations.

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}; \begin{bmatrix} 1 & x & x \\ x & 1 & x \\ x & x & 1 \end{bmatrix}; \begin{bmatrix} x & x & 1 \\ x & 1 & x \\ 1 & x & x \end{bmatrix}; \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}; \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix};$$

Fig. 1. The exemplary matrix types of structural elements used in erosion and dilatation operations (1-white point, x-insignificant)

In Fig. 1 the exemplary matrix types of structural elements used in process of erosion and dilatation of an image are shown.

The used type of the structural element has the basic influence on the result of the whole process of stave removal. In the case of removal of horizontal objects which thickness is few points maximum the element having few points in verse and one column is desirable.

The size of the structural elements which should be used in the erosion process depends on the thickness of a single line which is being removed. The rule that the structural element should have the verse number one more than the line thickness can be assumed.

So, the resolution of the input image has the basic importance for choosing the structural element. In the process of dilatation (the second stage of the stave removal method) the structural element should be smaller size than the one used in the erosion process. The examples of such elements are shown in Fig. 2.

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}; \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}; \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}; \begin{bmatrix} 11 \\ 11 \\ 11 \end{bmatrix}; \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix};$$

Fig. 2. Exemplary structural elements used in the dilatation process and erosion process during the stave removal

In the first stage of the realization of the method of staff line removal from musical notation the erosion operation has to remove horizontal lines. In Fig. 3 the example of musical notation is shown and in Fig. 4 the image after the erosion operation. As the structural element the matrix 9×1 was used. In Fig. 4 you can see that the staff line was removed but it caused (as a side effect) that the symbols are thinner. The aim of the erosion process realized in the second stage is to restore the previous sizes of the symbols. Fig. 5 shows the image after the erosion operation and further dilatation. The structural element was the matrix 3×2.

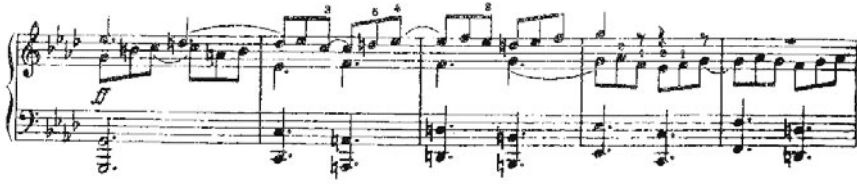


Fig. 3. The image of an exemplary sample of musical notation



Fig. 4. The image of a sample of musical notation after the erosion operation



Fig. 5. The image of a sample of musical notation after the stave removal (erosion operation and then dilatation one)

3 Preparing an Image to Notes Identification

The next stage of the process of recognition of musical notation is identification of the placement of musical symbols. The proper detection of the presence and correct position of the notes is basic for the whole process of recognition of musical notation. Variety of symbols in the musical notation can disturb in this stage, therefore all the signs and elements which are not notes should be removed. Similar to the staff removal the erosion and dilatation operations can be used here. The difference is in choosing of the structural elements. In erosion you should replace verses and columns (in the ratio to the structural element used in staff removal- in this case 1×9) The structural element in the dilatation operation should have twice less verses than columns (but the verses should be twice more than in the erosion operation- in this case 2×4). The result of these operation is shown in Fig. 6.

In the image after these operation the position of the notes, their distances and other relationships in the staff position can be identified.



Fig. 6. The sample of musical notation prepared to identify notes

4 Conclusions

Using of the morphology operation in the recognition of musical notation process let remove the staff simply and efficiently and prepare further stages of the recognition process. The results show that the properly prepared input image and the matching structural element remove efficiently the stave and do not make any important distortion at the same time (according to the stages of musical stages recognition).

References

1. Bainbridge, D., Bell, T.: An extensible optical music recognition system. *Australian Computer Science Communications* 18(1), 308–317 (1996)
2. Tadeusiewicz, R., Korohoda, P.: *Komputerowa analiza i przetwarzanie obrazów*. Wydawnictwo Fundacji Postępu Telekomunikacji, Kraków (1997)
3. Malina, W., Smiatacz, M.: *Metody cyfrowego przetwarzania obrazów*. Akademicka Oficyna Wydawnicza EXIT, Warszawa (2005)
4. Bainbridge, D.: Optical music recognition: A generalized approach. Paper read at Second New Zealand Computer Science Graduate Conference (1996)
5. Iwanowski, I.: *Metody morfologiczne w przetwarzaniu obrazów cyfrowych*. EXIT, Warszawa (2009)

GPU-Accelerated Object Tracking Using Particle Filtering and Appearance-Adaptive Models

Bogusław Rymut* and Bogdan Kwolek

Chair of Informatics and Control, Rzeszów University of Technology
W. Pola 2, 35-959 Rzeszów, Poland
e-mail: bkwolek@prz.edu.pl

Summary. In this work we present an object tracking algorithm running on GPU. The tracking is achieved by a particle filter using appearance-adaptive models. The main focus of our work is parallel computation of the particle weights. The tracker yields promising GPU/CPU speed-up. We demonstrate that the GPU implementation of the algorithm that runs with 256 particles is about 30 times faster than the CPU implementation. Practical implementation issues in the CUDA framework are discussed. The algorithm has been tested on freely available test sequences.

1 Introduction

Driven by the huge market for multimedia and games, graphics processors have evolved more quickly than CPUs, and currently outperform them not only in terms of processing power, but also in terms of memory performance. The increasing programmability and computational power of the graphics processing unit (GPU) provides great capability for acceleration of computer vision algorithms. The GPU computations are done in parallel and algorithms must work in multi-thread mode in order to exploit the computational power of the GPU [1], which is not a feature of many vision algorithms. Unlike traditional CPU-based programs, GPU-based programs have several limitations on how memory can be accessed. Thus, a majority of vision algorithms either cannot be implemented on current GPUs or can be implemented with considerable difficulties, in consequence leading to unsatisfactory speed-up.

The key to using the GPU for accelerating the computer vision algorithm is to view it as a streaming, data-parallel computer, and the computations in the form of SIMD, data-parallel kernels. GPU implementations should access constant memory efficiently, avoid shared memory bank conflicts, coalesce

* B. Rymut is presently a student, doing his MSc thesis on GPU-based object tracking.

global memory accesses, and overlap arithmetic with global memory latency. In general, the number of arithmetic operations must be high enough to effectively hide memory latency.

Visual tracking of objects of interest has received significant attention in the vision community. It is the key to the effective use of more advanced technologies, like human identification, event recognition, crowd analysis, etc. In the last decade a number of robust tracking strategies was proposed, which are able to tolerate changes in target appearance and track targets in complex scenes. One such successful approach is the particle filter [2, 3]. The most important property of the particle filter (PF) is its ability to handle complex, multi-modal (non-Gaussian) posterior distributions. Such distributions are approximated by a collection of the particles. Essentially, the number of particles required to adequately approximate the distribution grows exponentially with the dimensionality of the state space. PFs are computationally expensive as the number of particles needs to be large for precise results. Moreover, the observation models are often built on complex appearance models, and as the result the trackers have difficulties to operate with 25/30 frames per second.

Adaptive appearance models have demonstrated great effectiveness in object tracking. In [4], the appearance model is based on phase information derived from the image intensity. The appearance models [5, 6] consist of three components, namely W , S , F , where the W component represents the two-frame variations, the S component models temporally stable pixel intensities, and the F component is a fixed template of the target to prevent the model from drifting away. The particle filters built on adaptive appearance models algorithms produce good tracking results, but require considerable computational power. This motivated us to elaborate a GPU implementation of such an algorithm. Since the objects are represented as 2D arrays of pixels data, our algorithm takes advantages of GPU effectively.

The contribution of our work is an object tracking algorithm running on GPU. The tracking is achieved by a particle filter using appearance-adaptive models. The tracker yields promising GPU/CPU speed-up. We demonstrate that the GPU implementation of the algorithm that runs with 256 particles is about 30 times faster than the CPU implementation.

2 Object Tracking Using Appearance-Adaptive Models in Particle Filter

In this section we overview the particle filtering. The section explains also how the object undergoing tracking is modeled.

2.1 Particle Filtering

The particle filter simulates the behavior of the dynamical system. Each sample predicts future behavior of the system in a Monte-Carlo fashion, and

the samples that match the observed system behavior are kept, whereas ones that are unsuccessful in predicting tend to die out. The evolution of the state of the target as well as its measurement process is modeled by a set of (possibly non-linear) equations perturbed by (possibly non-Gaussian) i.i.d. noise:

$$\mathbf{x}_k = f_k(\mathbf{x}_{k-1}, \mathbf{v}_k) \quad (1)$$

$$\mathbf{z}_k = h_k(\mathbf{x}_k, \mathbf{n}_k) \quad (2)$$

where \mathbf{x}_k denotes the state of the target at discrete time k , \mathbf{v}_k is the process noise vector, \mathbf{z}_k is the measurement vector, and \mathbf{n}_k is the measurement noise vector. The aim is to estimate the distribution of the target state given all the previous measurements, that is, $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$, where $\mathbf{z}_{1:k-1} = \{\mathbf{z}_1, \dots, \mathbf{z}_{k-1}\}$. Given the initial distribution of the target, we can recursively predict the state of the target using:

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})d\mathbf{x}_{k-1} \quad (3)$$

If a new measurement becomes available, the state can be updated using Bayes' rule:

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})} \quad (4)$$

The complete tracking scheme, known as the recursive Bayesian filter first calculates the *a priori* density $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ using the system model and then evaluates *a posteriori* density $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ given the new measurement.

In the PF, the distribution $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$ is approximated by a set of M particles $\{\mathbf{x}_{k-1}^i\}_{i=1\dots M}$ and associated weights $\{w_{k-1}^i\}_{i=1\dots M}$ as follows: $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}) \approx \sum w_{k-1}^i \delta(\mathbf{x}_k - \mathbf{x}_{k-1}^i)$, where $w_k^i \propto w_{k-1}^i \frac{p(\mathbf{z}_k|\mathbf{x}_k^i)p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)}{q(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i, \mathbf{z}_k)}$, whereas $\sum w_{k-1}^i = 1$ and $\delta(\cdot)$ is the Kronecker delta function. The term $q(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i, \mathbf{z}_k)$ stands for an importance density, which is typically obtained by approximating $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_k)$ with a Gaussian distribution, or by using $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ like in CONDENSATION [2].

One of the practical difficulties that is associated with particle filters is degeneration of the particle population after a few iterations because weights of several particles are negligible, and, eventually, only a very small number of particles contributes to the posterior distribution. To mitigate this problem the resampling should be used in order to eliminate particles with low importance weights and multiply particles with high importance weights. Resampling can be carried out at every iteration or only when a substantial amount of degeneracy is observed [3].

The algorithm of the particle filter can be expressed in the pseudo-code:

1. For $i = 1, 2, \dots, M$ sample or propose particles using $p(\mathbf{x}_k|\mathbf{x}_{k-1})$
2. For $i = 1, 2, \dots, M$ calculate the weights, $\tilde{w}_k^i = w_{k-1}^i p(\mathbf{z}_k|\mathbf{x}_k^i)$
3. Normalize the weights w_k^i using \tilde{w}_k^i

4. Calculate the state estimates, $\hat{\mathbf{x}}_k = \sum_{i=1}^M w_k^i \mathbf{x}_k^i$
5. Resample $\{\mathbf{x}_k^i, w_k^i\}$ to get new set of particles $\{\mathbf{x}_k^j, w_k^j = 1/M\}$

2.2 Appearance-Adaptive Models

Our intensity-based appearance model consists of three components, namely, the W -component expressing the two-frame variations, the S -component characterizing the stable structure within all previous observations and F component representing a fixed initial template. The model $A_k = \{W_k, S_k, F_k\}$ represents the appearances existing in all observations up to time $k - 1$. It is a mixture of Gaussians [4] with centers $\{\mu_{k,l} \mid l = w, s, f\}$, their corresponding variances $\{\sigma_{k,l}^2 \mid l = w, s, f\}$ and mixing probabilities $\{m_{k,l} \mid l = w, s, f\}$.

Let $I(\mathbf{x}, k)$ denote the brightness value at the position $\mathbf{x} = (x, y)$ in an image \mathcal{I} that was acquired in time k . Let \mathcal{R} be a set of J locations $\{\mathbf{x}(j) \mid j = 1, 2, \dots, J\}$ defining a template. $Y_k(\mathcal{R})$ is a vector of the brightness values at locations $\mathbf{x}(j)$ in the template. The object likelihood is evaluated as follows:

$$p(\mathbf{z}_k | \mathbf{x}_k) = \prod_{j=1}^J \sum_{l=w,s,f} \frac{m_{k,l}(j)}{\sqrt{2\pi\sigma_{k,l}^2(j)}} \exp \left[-\frac{1}{2} \left(\frac{Y_k(j) - \mu_{k,l}(j)}{\sigma_{k,l}(j)} \right)^2 \right] \quad (5)$$

It uses a recursively updated appearance model, which depicts stable structures seen so far, two-frame variations as well as initial object appearance.

The update of the current appearance model A_k to A_{t+1} is done using the Expectation Maximization (EM) algorithm [7]. For a template $\hat{Y}_k(\mathcal{R})$, which is located in the image \mathcal{I} at position $\hat{\mathbf{x}}_k$, we evaluate the posterior contribution probabilities as follows:

$$o_{k,l}(j) = \frac{m_{k,l}(j)}{\sqrt{2\pi\sigma_{k,l}^2(j)}} \exp \left[-\frac{1}{2} \left(\frac{\hat{Y}_k(j) - \mu_{k,l}(j)}{\sigma_{k,l}(j)} \right)^2 \right] \quad (6)$$

where $l = w, s, f$ and $j = 1, 2, \dots, J$. The posterior contribution probabilities (with $\sum_k o_{k,l}(j) = 1$) are used in updating the mixing probabilities:

$$m_{k+1,l}(j) = \gamma o_{k,l}(j) + (1 - \gamma) m_{k,l}(j) \quad | \quad l = w, s, f \quad (7)$$

where γ is accommodation factor. Then, the first and the second-moment images are determined in the following manner:

$$M_{k+1}^{(1)}(j) = (1 - \gamma) M_k^{(1)}(j) + \gamma o_{k,s}(j) \hat{Y}_k(j) \quad (8)$$

$$M_{k+1}^{(2)}(j) = (1 - \gamma) M_k^{(2)}(j) + \gamma o_{k,s}(j) \hat{Y}_k^2(j) \quad (9)$$

In the last step the mixture centers and the variances are calculated as follows:

$$\mu_{k+1,s}(j) = \frac{M_{k+1}^{(1)}(j)}{m_{k+1,s}(j)}, \quad \sigma_{k+1,s}(j) = \sqrt{\frac{M_{k+1}^{(2)}(j)}{m_{k+1,s}(j)} - \mu_{k+1,s}^2(j)} \quad (10)$$

$$\mu_{k+1,w}(j) = \hat{Y}_k(j), \quad \sigma_{k+1,w}(j) = \sigma_{1,w}(j) \quad (11)$$

$$\mu_{k+1,f}(j) = \mu_{1,f}(j), \quad \sigma_{k+1,f}(j) = \sigma_{1,f}(j) \quad (12)$$

In order to initialize the model A_1 the initial moment images are set using the following formulas: $M_1^{(1)} = m_{1,s} Y_{t_0}(\mathcal{R})$ and $M_1^{(2)} = m_{1,s} (\sigma_{1,s}^2 + Y_{t_0}^2(\mathcal{R}))$.

3 Implementation of Object Tracking on GPU

At the beginning of this section we overview programming in CUDA framework. Afterwards we discuss implementation details of the algorithm on GPU.

3.1 Programming in CUDA

Compute Unified Device Architecture (CUDA) is a programming interface that employs the parallel architecture of NVIDIA GPUs for general purpose computing [8]. In CUDA, programs are expressed as kernels and GPU is viewed as a device that can carry out multiple concurrent threads. Threads are organized in two hierarchical levels, namely blocks, which are groups of threads executed on one of the GPU's multiprocessors, and grids, which are groups of blocks launched concurrently on the device, and which all execute the same kernel [1]. The memory requirements of a kernel determine how many threads can run concurrently on each multiprocessor. The threads in a block can share memory on a single multiprocessor. For a given kernel the block dimensions are chosen to optimize the utilization of the available computational resources. Warp is a group of threads executed physically in parallel in SIMD fashion. If the GPU processor must wait on one warp of threads, it simply starts executing work on a different one. Because registers are allocated to active threads, i.e. they stay allocated to the thread until it completes its execution, no swapping of registers and state takes place between GPU threads. In general, the more threads per block, the better the performance because the scheduler can better hide memory latencies. Large arithmetic calculations also contribute towards hiding the memory latency.

3.2 Implementation Details

Porting well known computer vision algorithms to GPUs is a challenging task. Creating efficient data structures for effective use of the GPU memory model is a challenging problem in itself [9]. In order to take the full advantages of the available resources one should make careful decisions according to the data layout, data exchange and synchronization. In order to decompose the algorithm onto GPU the data-parallel portions of the program should be identified and then separated as CUDA kernels.

The predicting of the particles, see pseudo-code in subsection 2.1, is done in a kernel, which uses the normally distributed random numbers. The random

numbers are generated in advance in two kernels. In the first one we generate pseudo-random numbers using the Mersenne Twister [10] kernel provided by the CUDA™ SDK. The second kernel employs the pseudo-random numbers to generate a set of normal random numbers. It uses Box Mueller transform based on trigonometric functions [11]. The random numbers are generated in 32 blocks and each block consists of 128 threads, where each thread generates two random numbers. In the kernel responsible for the prediction of the particles the position of each particle is calculated in a separate thread.

The calculation of the particle weights is done in two separate kernels. The first kernel performs the normalization of the pixels in the template to the unit variance, whereas the second one is executed after the first one and calculates the object likelihood (5). The size of the reference object template is 42×32 . In both kernels each thread processes one column of the template. For each particle the number of threads is equal to 32. The product in (5) is calculated using parallel reduction [8]. The results achieved by the first kernel are stored in the shared memory, and both kernels operate on textures.

The normalized weights w_k^i , and the state estimate $\hat{\mathbf{x}}_k$, see pseudo-code in subsection 2.1, are calculated with the use of the parallel reduction. The object state consists in the template location as well as its size. The admissible change of the template size between successive frames is ± 1 pixel. Given the object state, the update of the appearance model takes place. In the resampling step the multinomial algorithm [12] has been utilized. The vector of cumulative sums was extracted with the use of parallel reduction, whereas the random numbers were taken from the set that had been generated in advance.

When a new image becomes available, the algorithm scales down and scales up the input image. The aim of this operation is to provide the images from which we can extract object templates that are smaller/larger about one pixel with regard to the estimated template size. The images are scaled using the bilinear interpolation. In the discussed kernel, the number of blocks is equal to the number of columns of the input images, whereas the number of threads is equal to the number of rows.

4 Experimental Results

The experiments were conducted on a PC with 1 GB RAM, Intel Core 2 Quad, 2.66 GHz processor with NVIDIA GeForce 9800 GT graphics card. The graphics card has 14 stream multiprocessors, clocked at 1.5 GHz, each with 8 cores. It is equipped with 1024 MB RAM, 64 KB constant memory and 16 KB common memory. We implemented the algorithm in CUDA and compared the runtimes with its counterpart that was implemented in C/C++ and executed on the CPU. The CPU code was compiled with Visual Studio 2005 with the SSE2 (Streaming SIMD Extensions 2) option and O2 optimization turned on. Table 1 shows the running times and speed-up of the tracking algorithm both on CPU and GPU. The communication delays for transferring images

Table 1. Tracking times [ms] and speed-up obtained on CPU (Intel Core 2, 2.66 GHz) and on GPU (NVIDIA GeForce 9800 GT)

#particles	32	64	128	256	512
CPU	16.53	32.27	62.65	123.73	243.19
GPU	1.30	1.80	2.70	4.17	7.51
CPU/GPU	12.8	18.3	24.4	29.5	32.4

from CPU to GPU and vice versa have not been taken into account. The most time-consuming operation of the tracking algorithm is calculation of the likelihood function (5). This operation amounts to 0.82 of the whole processing time.

In the experiments we employed various number of particles, see Table 1. As we can observe, the algorithm achieves larger speed-up for larger number of particles. Figure 1 depicts some tracking results, which were obtained on gray images 1.

The dataset exhibits severe illumination conditions with partial shading. The template is a rectangular window initialized manually in the first frame. The initial template size for the Trellis70 dataset was set to 96×64 . The object tracking was performed in three dimensional space, i.e. we track the location of the template as well as its scale. The size of the reference frame was set to 42×32 . The maximal change of the template size between successive frames was constrained to ± 1 pixel. The example tracking results that are depicted on Fig. 1 were obtained using 32 particles.

**Fig. 1.** Face tracking using particle filter and adaptive appearance models. Frames #1, 50, 100, 150.

5 Conclusions

The adaptive appearance model-based particle filter is a robust algorithm for tracking objects. However, the computational cost of this algorithm is substantial. In this paper we presented our implementation of this algorithm on GPU. We explained how to design threads and memory structures for high performance. The result is a parallel algorithm that is easy to implement and yields promising GPU/CPU speed-up. The results showed that the GPU

¹ Trellis70 dataset is available at: <http://www.cs.toronto.edu/~dross/ivt/>

implementation of the algorithm running with 256 particles is about 30 times faster than the CPU implementation. Performance comparison on various CPU/GPU configurations of the particle filter is also presented.

References

1. Wasson, S.: Nvidia's GeForce 8800 graphics processor. Tech Report, November 8, PC Hardware Explored (2006)
2. Isard, M., Blake, A.: Condensation - conditional density propagation for visual tracking. *Int. J. of Computer Vision* 29, 5–28 (2006)
3. Doucet, A., Godsill, S., Andrieum, C.: On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing* 10, 197–208 (2000)
4. Jepson, A.D., Fleet, D.J., El-Maraghi, T.: Robust on-line appearance models for visual tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 25, 1296–1311 (2003)
5. Zhang, X., Hu, W., Maybank, S., Li, X., Zhu, M.: Sequential particle swarm optimization for visual tracking. In: *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Anchorage, AK, USA, pp. 1–8 (2008)
6. Kwolek, B.: Particle swarm optimization-based object tracking. *Fundamenta Informaticae* 95, 449–463 (2009)
7. Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B* 39, 1–38 (1977)
8. Nickolls, J., Buck, I., Garland, M., Skadron, K.: Scalable parallel programming with CUDA. *ACM Queue* 6, 40–53 (2008)
9. Lefohn, A.E., Sengupta, S., Kniss, J., Strzodka, R., Owens, J.D.: Glift: Generic, efficient, random-access GPU data structures. *ACM Transactions on Graphics* 25, 60–99 (2006)
10. Matsumoto, M., Nishimura, T.: Mersenne twister: a 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions on Modeling and Computer Simulation* 8, 3–30 (1998)
11. Box, G.E.P., Muller, M.E.: A note on the generation of random normal deviates. *The Annals of Mathematical Statistics* 29, 610–611 (1958)
12. Gordon, N.J., Salmond, D.J., Smith, A.F.M.: Novel approach to nonlinear/non-gaussian bayesian state estimation. In: *IEE Proc. part-F, Radar Signal Proc.*, vol. 140, pp. 107–113 (1993)

Application of Epipolar Rectification Algorithm in 3D Television

Jakub Stankowski and Krzysztof Klimaszewski

Poznan University of Technology Chair of Multimedia Telecommunicatians and Microelectronics, Polanka 3, 60-965 Poznań, Poland
e-mail: jstankowski@multimedia.edu.pl, kklima@et.put.poznan.pl

Summary. The paper describes a method for rectification of video in 3D television parallel multi-camera systems. Using the camera calibration data gathered, a new coordinate system is calculated, in which virtual camera positions and orientations are calculated. A rectifying perspective transform is calculated that performs transformation from the real camera coordinate system to a new system. Transformed images correspond to images from virtual rectified camera setup. The results obtained using the method described are verified by computation of depth maps using the rectified video.

1 Introduction

In contemporary 3D television systems, a video of a scene is captured by a number of synchronized cameras. In most cases, 3 or more cameras are used. When properly processed, information from those views can be used to provide a very convincing 3D sensation to the users. In most cases it is necessary to produce additional, virtual views to improve the 3D experience.

The virtual view generation from raw views is a complex process. To lighten the computational burden and to make it possible for a simple 3D television receiver to be able to produce virtual views, simplifying assumptions are made. The first one is that all cameras are placed along a straight line and have parallel optical axes. The second assumption is that 3D scene structure is known and given in a form of depth map. This approach is used in the MPEG group research activities regarding 3D television systems [1].

Efficient computation of depth maps is possible when all scene elements are projected to the images from consecutive cameras in a way that they lie on the same image line. This ensures that proper correspondences between camera views can be found in the same line of images from all cameras. In order to satisfy this requirement, all cameras need to be aligned so that their direction is identical and perpendicular to the straight line, along which the cameras are placed. This way of camera placement makes it easy to compute depth maps and, additionally, satisfies the first assumption specified

above. Unfortunately, it is practically impossible to position the cameras in such way. This is because cameras have different intrinsic and extrinsic parameters. Image sensors and optical system in the cameras, even from the same production batch, can differ in orientation and position significantly. It is also very difficult to ensure proper camera body alignment. A difference of 2 angular minutes results in a 1 pixel shift of the whole image.

Therefore images acquired by a camera system have to be transformed in a way that simulates the ideal camera position and orientation. Such preprocessing is called rectification.

The purpose of rectification is to produce artificial views that would be captured by hypothetical cameras with parallel camera axes, identical intrinsic parameters, and camera centers positioned along a straight line with all the image horizontal borders being parallel to the line of camera centers.

The simplest possible case, with 2 cameras in the system, has a solution that assures ideal rectification using perspective image transform. Ideal rectification is possible because two points in space always lie on a straight line, therefore it is only necessary to compensate intrinsic camera parameters differences and camera rotations. Algorithms for estimation of the camera parameters are given in [8, 9, 12], while the rectification algorithms for 2 camera case can be found in [2, 10, 11].

For 3 and more cameras there exist no theoretical solution for ideal rectification using image perspective transform. Therefore, the proposed algorithm, that is based on perspective transform, can only produce approximations of ideally rectified images. Despite of that, it performs very well, giving useful results.

Only little information about other approximate solutions can be found in literature, like those presented in [6, 7], however, according to authors knowledge, no implementation of this method is publicly available. Moreover, the article does not describe the whole process of performing rectification. In this article we present and describe in detail our original proposal of an algorithm of multiple camera system rectification algorithm.

2 Pinhole Camera Model

In our work we use the most popular camera model, which is the pinhole camera model [5]. In this model one assumes that projection of a scene on an image plane can be described by the following equation:

$$z \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{A} [\mathbf{R}|\mathbf{T}] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \quad (1)$$

where: \mathbf{A} is $[3 \times 3]$ intrinsic matrix, containing information about camera principal point and focal length. It also contains information about physical image

cell proportions and skew of the consecutive image lines, \mathbf{R} is $[3 \times 3]$ rotation matrix, carrying information about camera orientation, \mathbf{T} is $[3 \times 1]$ translation vector, containing information about camera location, X, Y, Z are real world 3D coordinates of a point, u, v are image coordinates, z is scaling coefficient.

To model the distortions caused by the real camera optics, additional distortion parameters are introduced.

3 Proposed Rectification Algorithm

The proposed algorithm was developed for the parallel multi-camera system, like the one used in Poznan University of Technology [13]. The algorithm makes use of procedures implemented in OpenCV library [3].

3.1 Camera Calibration

The first step in the solution proposed is to calibrate all cameras. We need to calculate values of intrinsic matrix \mathbf{A} , distortion coefficients \mathbf{D} , rotation matrix \mathbf{R} and translation matrix \mathbf{T} for every camera. One of the possible ways to obtain those parameters is to use an algorithm proposed in [4] that makes use of images of a known chessboard pattern placed in front of the camera and is available as a part of OpenCV [3] library.

First we independently calibrate every camera. This operation is performed in order to obtain accurate intrinsic matrix and distortion coefficients for every camera. We use two-dimensional, high-contrast chessboard calibration pattern. The size of this pattern allows us to fill most frame area by the pattern image and improves distortion estimation, especially near the borders of image. The OpenCV corner detection algorithm is used to find corners coordinates on all captured images with sub-pixel precision. Corners coordinates, averaged over 100 frames of the same pattern position, were used to perform camera calibration. This was done using algorithm proposed in [4]. Calculated intrinsic matrices and distortion coefficients are saved and used in further calculations.

The next step is to find translation and rotation matrices. This operation is performed by capturing calibration pattern simultaneously by all cameras in the system. We use the same calibration pattern, capturing procedure and corners coordinates estimation method as in previously. Calculated coordinates and already known cameras intrinsic parameters are used to find extrinsic cameras parameters. The extrinsic parameters (\mathbf{R} , \mathbf{T}) are calculated with respect to chessboard pattern using the OpenCV library. Calculated translation \mathbf{T} vectors and rotation \mathbf{R} matrices are the base for next steps.

3.2 Coordinate Systems Transposition

After extrinsic parameters calculation we proceed to estimate relations between all cameras in a multi-view system. Translation vectors \mathbf{T} and rotation

matrices \mathbf{R} were calculated in every camera own coordinate system, so the important operation is to present these parameters into one coordinate system. In the first step of finding inter-camera relations, we use first camera (camera 0) coordinate system as a common base for the whole camera system. We calculate relative rotation (\mathbf{R}_{rel}) and relative translation (\mathbf{T}_{rel}) between camera 0 and other cameras:

$$\begin{aligned} \mathbf{R}_{rel}(0 \rightarrow i) &= \mathbf{R}(0) \cdot \mathbf{R}^{-1}(i), \\ \mathbf{T}_{rel}(0 \rightarrow i) &= \mathbf{R}_{rel}(0 \rightarrow i) \cdot \mathbf{T}(i) - \mathbf{T}(0). \end{aligned} \quad (2)$$

where $\mathbf{R}_{rel}(0 \rightarrow i)$ is rotation matrix between camera 0 and cam. i , $\mathbf{R}(i)$ is camera i rotation matrix, $\mathbf{T}_{rel}(0 \rightarrow i)$ is translation matrix between camera 0 and camera i , $\mathbf{T}(i)$ is cam. i translation matrix. Those calculated inter-camera relations are still related to camera 0 position and direction, so the next important operation is to find a coordinate system that does not promote any camera. The new coordinate system has to be independent of any single camera coordinate system. We find this system by using linear regression and try to estimate a line nearest to center points of all cameras. This line forms the x-axis of a new coordinate system. Linear regression is calculated separately for $y(x)$ and $z(x)$ relations between cameras center points coordinates. As a result of the regression analysis, we get a_y, b_y, a_z and b_z coefficients:

$$\begin{aligned} y(x) &= a_y \cdot x + b_y, \\ z(x) &= a_z \cdot x + b_z. \end{aligned} \quad (3)$$

After we estimate linear regression, we calculate translation and rotation that transform camera 0 coordinate system to the new system. The y- and z-intercepts (b_y, b_z) of linear regression are used to create translation correction (\mathbf{T}_c) vector, which corresponds to the origin points difference of the old and new system:

$$\mathbf{T}_c = \begin{bmatrix} 0 \\ -b_y \\ -b_z \end{bmatrix}. \quad (4)$$

where \mathbf{T}_c is translation correction vector. Next step is to estimate rotation between camera 0 coordinate system and the new coordinate system. We use slopes of the lines (a_y, a_z) to calculate rotation correction matrix (\mathbf{R}_c). Angle of rotation around z axis (α_z) is calculated using a_y coefficient and rotation around y axis (α_y) is calculated using a_z coefficients:

$$\begin{aligned} \alpha_z &= \tan^{-1}(a_y), \\ \alpha_y &= \tan^{-1}(a_z). \end{aligned} \quad (5)$$

Previously obtained rotation angles are then converted into rotation matrices. The z-axis rotation matrix (\mathbf{R}_z) is calculated using α_z rotation angle, and y-axis matrix (\mathbf{R}_y) is calculated using α_y rotation angle (see Fig. [1](#)):

$$\mathbf{R}_y = \begin{bmatrix} \cos(\alpha_y) & 0 & \sin(\alpha_y) \\ 0 & 1 & 0 \\ -\sin(\alpha_y) & 0 & \cos(\alpha_y) \end{bmatrix}, \mathbf{R}_z = \begin{bmatrix} \cos(\alpha_z) - \sin(\alpha_z) & 0 \\ \sin(\alpha_z) & \cos(\alpha_z) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (6)$$

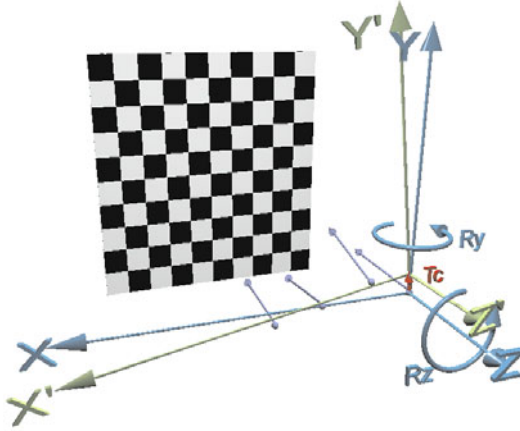


Fig. 1. Transformation between coordinate systems. X, Y, Z is camera 0 coordinate system. X', Y', Z' is independent coordinate system.

Rotation between camera 0 coordinate system and new coordinate system is calculated as a rotation correction matrix (\mathbf{R}_c) as a product of \mathbf{R}_y and \mathbf{R}_z matrices:

$$\mathbf{R}_c = \mathbf{R}_z \cdot \mathbf{R}_y. \quad (7)$$

Previously calculated relative translation vectors have to be converted to the new coordinate system. This allows us to find cameras coordinates in the new coordinate system:

$$\mathbf{T}_n(i) = \mathbf{R}_c \cdot [T_{rel}(0 \rightarrow i) + \mathbf{T}_c]. \quad (8)$$

where $\mathbf{T}_n(i)$ is camera i coordinate in new coordinate system, $T_{rel}(0 \rightarrow i)$ is translation matrix between camera 0 and camera i , R_c is rotation correction matrix, T_c is translation correction matrix, i is camera number.

The last estimated parameter is an average rotation matrix \mathbf{R}_a , calculated by averaging relative rotation matrices $R_{rel}(0 \rightarrow i)$ for all cameras. This parameter is used to find the difference between camera 0 rotation and average rotation of all cameras. R_a parameter will be used to eliminate camera 0 rotation influence in target camera parameters calculation.

3.3 Final Camera Parameters Calculation

Previously calculated real cameras parameters and cameras relations are now used to find target cameras parameters. These parameters correspond to parameters of ideal multi-view system. Target intrinsic matrix \mathbf{A}_t is calculated by averaging all cameras intrinsic matrices. This forces all cameras to have the same focal length and the same position of the principal point. We also

assume that target cameras are distortion free, so set the distortion coefficients in target system to zero. Target translation and rotation matrices were calculated using known translation and rotation matrices and correction matrices (\mathbf{R}_c , \mathbf{T}_c , and \mathbf{R}_a). Target parameters for camera 0 are calculated as follows:

$$\begin{aligned}\mathbf{T}_t(0) &= \mathbf{R}_c \cdot [\mathbf{T}(0) + \mathbf{T}_c], \\ \mathbf{R}_t(0) &= \mathbf{R}_a^{-1} \cdot [\mathbf{R}_c \cdot \mathbf{R}(0)].\end{aligned}\quad (9)$$

where $\mathbf{T}_t(0)$ is target translation matrix for first camera, $\mathbf{T}(0)$ is first camera translation matrix, $\mathbf{R}_t(0)$ is target rotation matrix for first camera, $\mathbf{R}(0)$ is first camera rotation matrix, \mathbf{T}_c is translation correction matrix, \mathbf{R}_c is rotation correction matrix, \mathbf{R}_a is average rotation matrix. Target rotation matrices for all cameras are the same as for the camera 0. This makes cameras optical axes parallel. Target translation matrices for other cameras are calculated by separating their new optical centres positions along x axis by a defined step, calculated as an average distance between neighbouring cameras (d_a). This operation was performed by modifying x-axis value in target translation matrix:

$$\mathbf{T}_t(i) = \mathbf{T}_t(0) + \begin{bmatrix} d_a \cdot i \\ 0 \\ 0 \end{bmatrix}, \quad (10)$$

where $\mathbf{T}_t(i)$ is target translation matrix for i camera, $\mathbf{T}_t(0)$ is target translation matrix for first camera, d_a is average distance between cameras, i is camera number.

3.4 Rectifying Transform Calculation

Previously calculated ideal multi-view system cameras parameters are now used to estimate rectification parameters. We select four points from corners of calibration pattern, and correct their coordinates to remove distortion influences. We also use the target cameras parameters and the equation (1) to calculate projection of points that correspond to the corners of calibration pattern. This allows us to obtain corners coordinates in ideal multi-view system. Having two groups of four points (real and target), we are able to find a perspective transformation \mathbf{P}_t that performs mapping between these points. The final part of the described solution is image transformation. First, the distortions are removed from the image with the use of known intrinsic matrix \mathbf{A} and distortion vector \mathbf{D} . Then, a perspective transformation \mathbf{P}_t is applied to all images.

4 Conclusions

In the application this method is targetted to, the quality of rectification can be judged by calculating depth maps. Resulting depth map can be compared with the one calculated with original images. Result of such test with



Fig. 2. Image of the scene (left), depth map estimated for test sequence without rectification (top right) where significant artefacts can be seen and after rectification with proposed algorithm (bottom right) with very limited amount of artefacts

a depth map estimation before and after performing rectification using algorithm proposed can be seen on Fig. 2. On the depth map generated with rectified images, object boundaries are clearly visible, and correspond to the boundaries of the respective objects in video frame. Disparity values are calculated more accurately.

The algorithm is proved to be an effective way of performing rectification in multi-camera system. Its main application is to produce images that satisfy the requirements of depth estimation algorithms and simplified view synthesis algorithms. The algorithm described here was used to prepare 4 test sequences contributed to MPEG works [14]. These sequences are regarded as very good and are used by many research groups in the world for experiments in 3D video field.

The use of proposed algorithm increases overall quality of generated depth map compared to the case with unrectified images. It is an universal and straightforward solution for parallel multi-camera setups.

Acknowledgment

This work was supported by the public funds as a research project.

References

1. Description of Exploration Experiments in 3D Video Coding. ISO/IEC JTC1/SC29/WG11, MPEG 2010/W11274, Dresden, Germany (2010)
2. Fusiello, A., Trucco, E., Verri, A.: A compact algorithm for rectification of stereo pairs. In: *Machine Vision and Applications*, vol. 12(1), Springer, Heidelberg (2000)
3. OpenCV homepage, <http://sourceforge.net/projects/opencvlibrary/>
4. Zhang, Z.: A Flexible New Technique for Camera Calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(11), 1330–1334 (2000)
5. Cyganek, B., Siebert, J.P.: *An Introduction to 3D Computer Vision Techniques and Algorithms*. John Wiley and Sons, Chichester (2009)
6. Yun-Suk, K., Yo-Sung, H.: Geometrical Compensation for Multi-view Video in Multiple Camera Array. In: 50th International Symposium ELMAR 2008, Zadar, Croatia, September 10-12 (2008)
7. Yun-Suk, K., Cheon, L., Yo-Sung, H.: An Efficient Rectification Algorithm for Multi-view Images in Parallel Camera Array. In: *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, Istanbul, May 28-30 (2008)
8. Zhang, Z.: Flexible Camera Calibration By Viewing a Plane From Unknown Orientations. In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 1, pp. 666–673 (1999)
9. Tsai, R.: A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation* 3(4), 323–344 (1987)
10. Hartley, R.: Theory and Practice of Projective Rectification. *International Journal of Computer Vision* 35(2), 115–127 (1999)
11. Loop, C., Zhang, Z.: Computing rectifying homographies for stereo vision. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 125–131 (1999)
12. Nowakowski, A., Tomaszewski, M., Skarbek, W.: Modelowanie Nieliniowości Optycznych w Obrazie Cyfrowym Technika Iteracji Punktów Centralnych. In: *V Sympozjum Naukowe, Techniki Przetwarzania Obrazu*, Serock (2006) (in Polish)
13. Domański, M., Klimaszewski, K., Konieczny, J., Kurc, M., Łuczak, A., Stankiewicz, O., Wegner, K.: An Experimental Free-view Television System. In: Choraś, R., Zabłudowski, A. (eds.) *Image Processing and Communications Challenges*, pp. 169–176. Academy Publishing House EXIT, Warsaw (2009)
14. Domański, M., Grajek, T., Klimaszewski, K., Kurc, M., Stankiewicz, O., Stankowski, J., Wegner, K.: *Poznan Multiview Video Test Sequences and Camera Parameters*. ISO/IEC JTC1/SC29/WG11, MPEG 2009/M17050, Xian, China (2009)

Crack Detection on Asphalt Surface Image Using Local Minimum Analysis

Adam Marchewka

Institute of Telecommunication, University of Technology and Life Sciences,
Kaliskiego 7, 85-791 Bydgoszcz
e-mail: adimar@utp.edu.pl

Summary. In this paper a new method of cracks detection is introduced. The proposed algorithm is applied to detect the cracks in the pavement image. Local minimum and linear relation between them was proposed. The proposed method is classify into two stages: linear local minimum and verification of detecting of pavement cracking. This method is fast although is complex. Additionally, the proposed method eliminates slight and strong variations like irregularly illuminated conditions, shading and road signs painted on pavement surface.

1 Introduction

The inspection of the concrete structures is a major part of roads management. For many years manual observation of the road surfaces has been the most common method for evaluating road surface distress around the world [1, 2, 4, 5, 8].

The general approach to the defect detection in pavement surface distress is to find a "homogeneous" feature of "no defect" textures and to detect the differences caused by presence of a defect. Major supposes are the cracked regions darker than other regions of inspected images and a crack is a continuous region. Usually the first step is to improve the contrast of inspected image, being less dependent on the illumination condition and the type of textures in road surface. The second step is camera calibration.

Over the past several decades, a number of approaches for automatic pavement cracking detection have been proposed. One of them is a ACP system [3, 12] to detect horizontal and vertical cracks, crack lengths, and severity. This is a block-based method. For all block (48×48) it was calculated vertical and horizontal projection histogram. In [9, 10] wavelet transform to crash detection surface was presented. There was proposed method consisted of three stages: pre-processing, wavelet processing, post-processing. Different approach to detect pavement destroy is proposed in [13]. There are three

steps: subtraction pre-processings, line emphasis pre-processing and thresholding processing. Modification of this method is shown in [7]. Morphology to the pavement crack detection and the median-filter algorithm with four structural elements we found in [6]. Authors proposed method to avoid noises more ideally and introduced the structural element to median-filter of four shapes. The similar algorithm was shown in [11]. Dilatation and erosion operators were also adapted. The difference between both algorithms is pre-processing phase. Here the pre-processing is based on a histogram equalization (to be less dependent on the illumination condition and the type of textures in road surface) and two thresholds during binarization process.

2 Proposed Algorithm

In digital image cracks are characterized as basic features:

- the pavement distress is dark. Its means than the region of asphalt where are cracks is darker than the other regions;
- pounding pavement surface segment is linear;
- not all dark points build a distress.

From this assumption arises the fact that we can not use simple algorithm like thresholding to separate cracks. The key point is to study over methods which count all foregoing assumption.

In this paper was proposed a new method of cracks detection which does not use pre-processing stage. The algorithm is applied to detect cracks on the pavement image. I propose to find local minimums and to describe the linear relations between them. This method consists of two stages: local minimums finding and verification.

2.1 Local Minimum Searching

In the first stage proposed Linear Local Minimum algorithm searches for local minimums L_x , L_y , which are described by formula (1) and (2) respectively:

$$L_x[x, y] = \begin{cases} 1 & \forall_{j \in \{1, 2, \dots, N\}} : I[x, y] = \min\{I[x, j] | x \in \{1, 2, \dots, M\}\} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$L_y[x, y] = \begin{cases} 1 & \forall_{j \in \{1, 2, \dots, M\}} : I[x, y] = \min\{I[j, y] | y \in \{1, 2, \dots, N\}\} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

This way we obtain two matrix which are added to produce matrix $L[x, y]$, described by eq. (3)

$$L[x, y] = L_x[x, y] + L_y[x, y] \quad (3)$$

where: $L : \{1, 2, \dots, M\} \times \{1, 2, \dots, N\} \rightarrow \{0, 1, 2\}$, N and M are width and height of image.

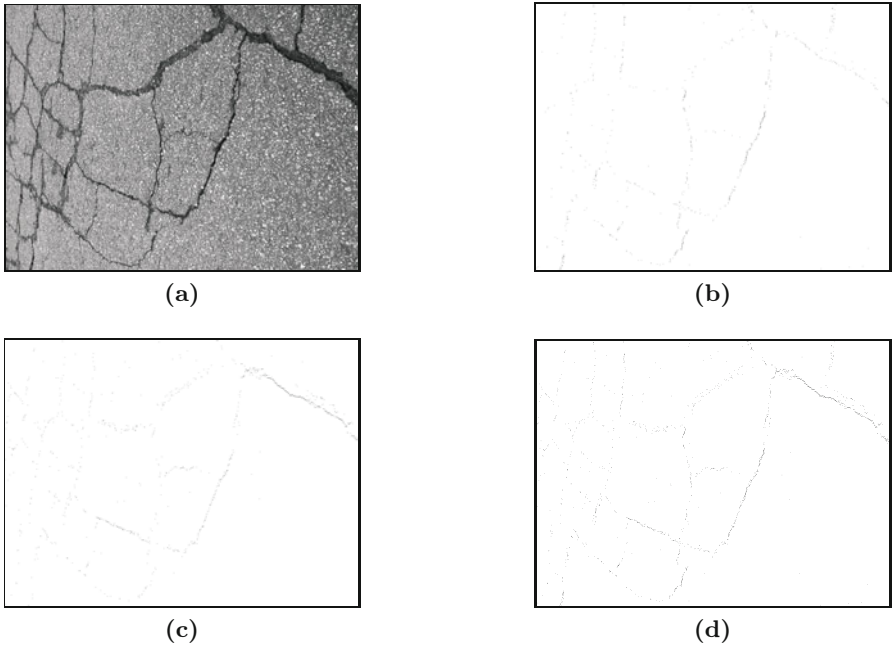


Fig. 1. Image I and computed arrays L_x (b), L_x (c) and L_T (d)

We can note that the value of $L[x, y]$ is interpreted as:

- $L[x, y] = 0$ - in $[x, y]$ point of image I is a not local minimum;
- $L[x, y] = 1$ - $[x, y]$ is a point of image I where is vertical or horizontal minimum;
- $L[x, y] = 2$ - $[x, y]$ is a point of image I where we found a local minimum in two directions;
- maximum amount of non zero elements in matrix $L[x, y]$ is equal to $N+M$, and minimum is equal to $\sqrt{N^2 + M^2}$.

The array of labels $L[x, y]$ implies only the most probable points fall into cracks. In other words, the non zero elements of $L[x, y]$ are only the candidate points, which increase the probability of place being the crack is. Additionally, the number of non zero elements of $L[x, y]$ are not impacting to the results. That's why the binarization process is calculated as:

$$L_T[x, y] = \begin{cases} 1 & \text{if } L[x, y] > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

2.2 Verification Process

In this stage we eliminate "false" points in L_T matrix and classify other as crack or not. We defined line segment between two points $L_T[x_1, y_1]$ and $L_T[x_2, y_2]$:

$$l_{1,2} : (y - y_1)(x_2 - x_1) - (y_2 - y_1)(x - x_1) = 0 \quad \forall_{x \in \langle x_1, x_2 \rangle, y \in \langle y_1, y_2 \rangle} \quad (5)$$

The line $l_{1,2}$ is classified as the crack when the following condition is met

$$M_e \geq \frac{1}{l_{1,2}(S)} \sum_{a,b \in S} (I[a, b] - Me) - \tau \quad (6)$$

where $M_e = \text{mean}(I[x_1, y_1], I[x_2, y_2])$ and S is a set of points of line segment $l_{1,2}$. This condition is checked for all pairs of nonzero point from array L_T .

To decrease the number of calculations we defined new condition: the d between two point $L_T[x_1, y_1]$ and $L_T[x_2, y_2]$

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (7)$$

must be shorter than established value r .

The proposed algorithm of verification process is show in Algorithm 3.

Algorithm 3. verification algorithm

Input: $I[x, y]$ - original image, $L_T[x, y]$ - binary local minimum array

Output: $B[x, y]$ - array of labels

```

for all  $L[x, y] \neq 0$  do
  if  $|L[x_1, y_1]L[x_2, y_2]| < r$  where  $x_1 \neq x_2$  or  $y_1 \neq y_2$  then
    if  $L_T[a, b] \in S$  then
      if  $\text{median}(I[x_1, y_1], I[x_2, y_2]) \geq \frac{1}{l(S)} \sum_{a,b \in S} I[a, b] - \tau$  then
         $B[a, b] = 1$ 
      end if
    end if
  end if
end if
end for
    
```

3 Experimental Results

Applying proposed method on sample images, the results were obtained (see Fig. 2-6). When we increase r we go up with probability of finding all "alligators" cracks, but too big r increaser false detection ratio.

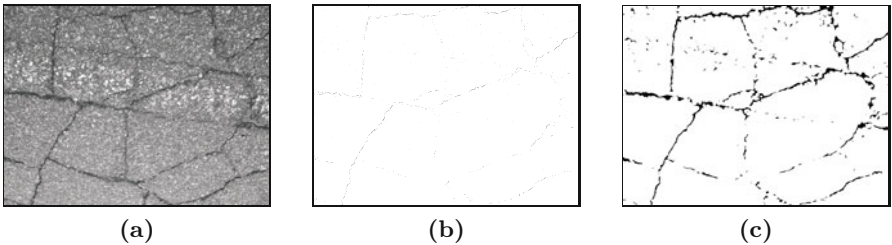


Fig. 2. Test Image #1, (a) array of labels L , (b) binary cracks mask B

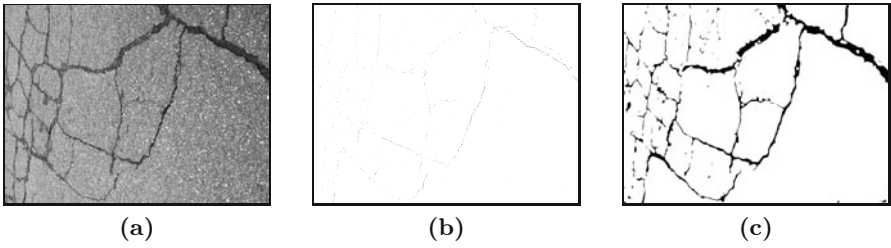


Fig. 3. Test Image #2, (a) array of labels L , (b) binary cracks mask B

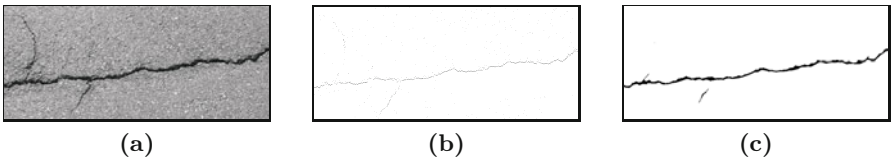


Fig. 4. Test Image #3, (a) array of labels L , (b) binary cracks mask B

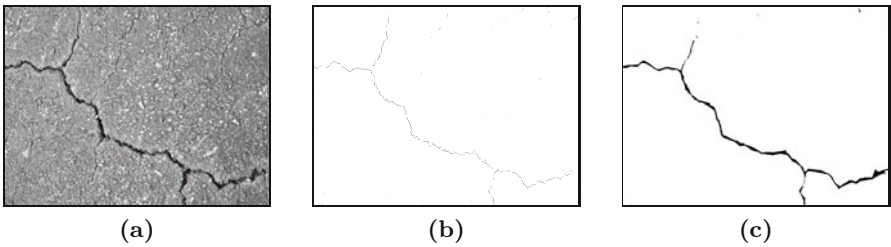


Fig. 5. Test Image #4, (a) array of labels L , (b) binary cracks mask B

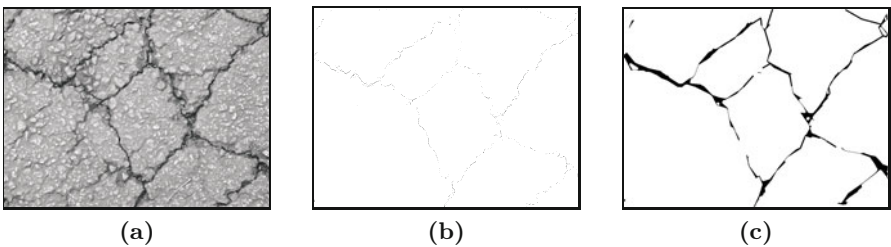


Fig. 6. Test Image #5, (a) array of labels L , (b) binary cracks mask B

Additionally, the white and yellow lines are not classified as cracks - what is the biggest advantage. The results are shown in Table II. This approach gives 10% of false positive, 19% of false negative in the summary results.

Table 1. Testing Result on Asphalt Road Images

	Method [11]		Method [6]		Proposed method	
	true	false	true	false	true	false
Non-Cracked Images	87%	13%	89%	9%	85%	10%
Cracked Images	79%	21%	78%	22%	81%	19%

4 Conclusions

The proposed method is detecting a complex linear local minimum of pavement cracking. This method is fast although is complex. Additionally proposed method eliminates slight and strong variations like irregularly illuminated conditions, shadings and road signs painted on pavement surface.

References

1. ASTM Standard practice for roads and parking lots. Pavement Condition Index surveys ASTM designation D 6433-99 (1999)
2. Austroads: A Guide to the Visual assessment of pavement condition, Sydney (1987)
3. Chan, P., Rao, L.L., Lytton, L.R.: Development of Image Algorithms for Automated Pavement Distress Evaluation System. FHWA Report TX-92/1189-2F. TX: Texas Transportation Institute, Texas A and M University (1992)
4. FHWA Distress identification manual for the long-term pavement performance project. FHWA-RD-03-031 (2003)
5. GDDP - BSSD System oceny stanu nawierzchni SOSN - wytyczne stosowania, Warszawa (2002)
6. Maode, Y., Shaobo, B., Kun, X., Yuyao, H.: Pavement Crack Detection and Analysis for High-grade Highway. In: Eighth International Conference on Electronic Measurement and Instruments, ICEMI (2007)
7. Marchewka, A., Miciak, M.: Subtract-filtering Pre-processing for Cracks Detection. In: Choraś, R.S., Zabłudowski, A. (eds.) Image Processing & Communications Challenges, pp. 225–230. Academy Publishing House EXIT, Warsaw (2009)
8. Marchewka, A.: Location Of Pavement Surface Distress Using Digital Processing - A Survey. Image Processing & Communications (2009)
9. Subirats, P., Fabre, O., Dumoulin, J., Legeay, V., Barba, D.: A Combined Wavelet-Based Image Processing Method for Emergent Crack Detection on Pavement Surface Images. In: EUSIPCO, Vienna, Austria (2004)
10. Subirats, P., Dumoulin, J., Legeay, V., Barba, D.: Automation of Pavement Surface Crack Detection Using The Continuous Wavelet Transform. In: ICIP, pp. 3037–3040 (2006)
11. Sy, N.T., Avila, M., Begot, S., Bardet, J.C.: Detection of Defects in Road Surface by a Vision System. In: The 14th IEEE Mediterranean, Electrotechnical Conference, MELECON 2008, pp. 847–851 (2008)

12. Teomete, E., Amin, V.R., Ceylan, H., Smadi, O.: Digital Image Processing for Pavement Distress Analyses. In: Proceedings of the 2005 Mid-Continent Transportation Research Symposium, Ames, Iowa (2005)
13. Yusuke, F., Yoshihiro, M., Yoshihiko, H.: A Method for Crack Detection on a Concrete Structure. In: ICPR 2006: Proceedings of the 18th International Conference on Pattern Recognition, pp. 901–904. IEEE Computer Society, Washington (2006)
14. Xu, B., Huang, Y.: Automatic Inspection of Pavement Cracking Distress. *Journal of Electronic Imaging* 15 (2006)

Eye Tracking System for Human Computer Interaction

Mscisław Śrutek and Łukasz Matuszak

Institute of Telecommunications, University of Technology and Life Sciences
e-mail: michus@utp.edu.pl

Summary. Algorithm of eyes and iris detection for mouse cursor control is proposed. Non-invasive method analyzes images from usb camera. After eye detection, coordinates of iris position are determined. After short calibration, iris movement is converted into cursor position. Application based on OpenCV library has been created.

1 Introduction

Computer mouse is a basic device used for computer control but sometimes hands free interaction with computer is needed (e.g. physically handicapped persons). Research concerning alternative methods of controlling the computer has been growing. Those ways of human computer interaction (HCI) are based on face gestures and head rotation information [5], brainwaves analysis or movement of infrared illuminated eyes (this is controversial method) [10].

The aim of this paper was to develop method to detect eyes, determine iris positions and create an application, which accordingly to their positions will control mouse cursor. For this project OpenCV (Open source Computer Vision library) developed by Intel Corporation, were used. It allows high level functions for computer vision and image processing.

Most eye recognition and eye detecting systems uses infrared waves which bounces from the retina and are caught by infrared sensitive cameras to ease the pupil search algorithms (as shown in Fig. 2). In this paper a method to do this using usb camera and PC was proposed.

Since eye moves in relatively small range there are two possibilities to provide effective system:

- High resolution camera or camera with good optical zoom.
- Place camera as close to an eye as possible.

The disadvantages of the first possibility are that system will have larger image to process which could cause system to work slowly. On the other hand second solution is less comfortable because the camera cannot be placed on



Fig. 1. Camera on glasses holder

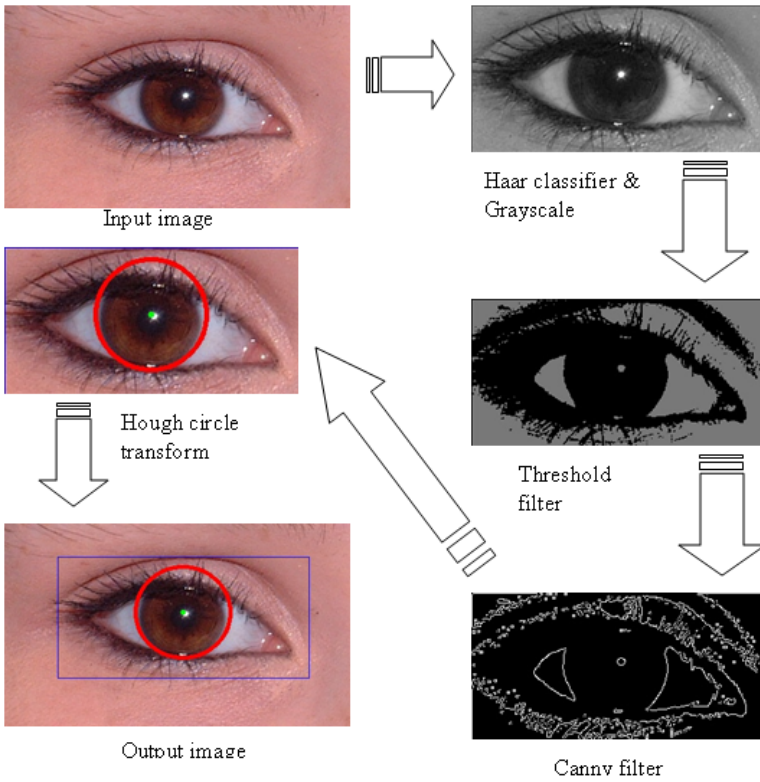


Fig. 2. Algorithm diagrams

your desk. After tests with varied usb cameras, second one (placing camera lens on the glasses holder) was used. Constructed device is shown in Fig. 1.

Algorithm

The general overview of the proposed methodology is shown in Fig. 2.

2 Object Detection

The algorithm is divided into few steps. At the first step, algorithm must search for an eye (if it is present on the captured image). For these Haar-like features classifier is used. It uses AdaBoost algorithm. The description of AdaBoost algorithm is shown below as pseudo-code. The more detailed description is presented in [2]. On the input there's train set provided $(x_1, y_1), \dots, (x_m, y_m)$, where x_i belongs to the space of patterns x and y_i is labeling pattern.

AdaBoost algorithm chooses weak classifiers $h_t(x) = \{-1v + 1\}$ in series of rounds $t = 1, \dots, T$. Output classifier $H(x)$ is line combination of weak classifiers h_t with weights α_t assigned in the process of learning.

Algorithm 4. AdaBoost algorithm

Input: $(x_1, y_1), \dots, (x_m, y_m)$, where $x_i \in X, y_i \in Y = \{-1, +1\}$

Output: $H(x) = \text{sgn}\left(\sum_{t=1}^T \alpha_t h_t(x)\right)$

for $i = 1$ to m **do**

initialize $D_1(i) = \frac{1}{m}$.

end for

for $t = 1$ to T **do**

find $h_t = \arg \min_{h \in \mathcal{X}} \varepsilon = \sum_{i: y_i \neq h(x_i)} D_t(i)$

if $\varepsilon_t \geq \frac{1}{2}$ **then**

stop

end if

where ε_t is an error of selected classifier h_t

calculate: $\alpha_t = \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$.

for $i = 1$ to m **do**

calculate: $D_{(t+1)}(i) = \frac{D_t(i)e^{-y_i \alpha_t h_t(x_i)}}{Z_t}$

end for

end for

Important issue while training classifier is to use proper images to train the system. Fig. 3 and Fig. 4 shows example of images that were used (set of samples were created from larger images).

Total number of the used positive images (eye images) was **396** and total number of negative images (images without eyes) was **287**. Training process was divided into **15 stages** and in each stage there were two splits (weak

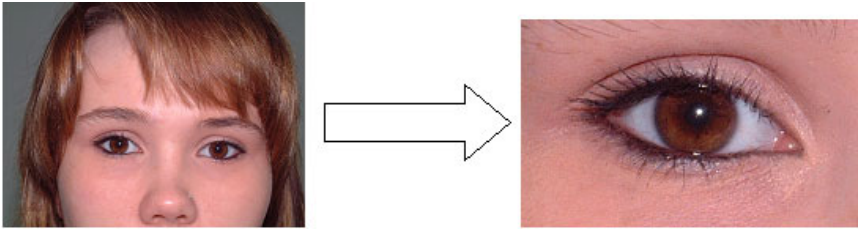


Fig. 3. Positive training images



Fig. 4. Negative training images

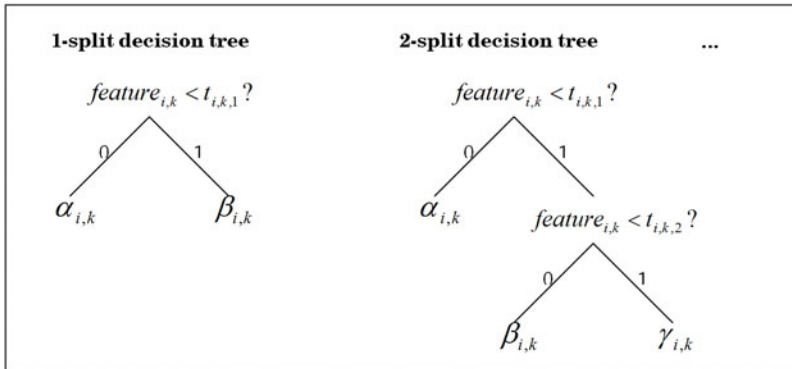


Fig. 5. Weak classifiers decision trees

classifiers [1]) calculated. Example decision trees (weak classifiers) are shown in Fig. 5.

$t_{i,k}$ and values of each leafs are calculated by *L. Brieman CARTTM* [3]. Total training took about 22 hours to complete (PC: AMD Turion 2×2GHz, 4GB DDR2). Final stage number was 13 since trainer reached maximum false alarm rate.

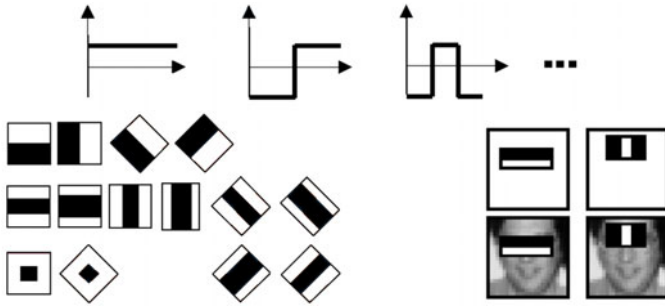


Fig. 6. Pool of features used in OpenCV implementation

The classifier is then used as Haar-Like object detector. It is called Haar-like, because the features are similar to basic functions in Haar wavelets. Such classifiers are showed in Fig. 6.

In this paper extended set of Haar-like features were used [8]. Features can be scaled and there can be about 118 000 features for 24×24 window. Below is the implementation of how each feature is calculated.

$$feature_{i,k} = w_{i,k,1} * RectSum_{i,k,white+black}(I) + w_{i,k,2} * RectSum_{i,black}(I) \tag{1}$$

Weights are compensated:

$$w_{i,k,1} * Area_{i,k,white+black} + w_{i,k,2} * Area_{i,k,black} = 0 \tag{2}$$

3 Pupil Center Detection

Aside from detecting the eye the more important part of the algorithm is to detect center of the pupil. The idea of the algorithm comes from [9]. Hough circle transform [6] is used to determine the parameters of a circle when a number of points that fall on the perimeter are known. A circle with radius R and center (a, b) can be described with the parametric equations.

$$x = a + Rcos(\Theta) \ ; \ y = b + Rsin(\Theta) \tag{3}$$

When the angle Θ sweeps through the full 360 degree range the points (x, y) trace the perimeter of a circle.

If an image contains many points, some of which fall on perimeters of circles, then the job of the search program is to find parameter triplets (a, b, R) to describe each circle. The fact that the parameter space is 3D makes a direct implementation of the Hough technique more expensive in computer memory and time. If the circles in an image are of known radius R , then the search can be reduced to 2D. The objective is to find the (a, b) coordinates of the centers. The locus of (a, b) points in the parameter space fall on a circle of radius R centered at (x, y) . The true center point will be common to all

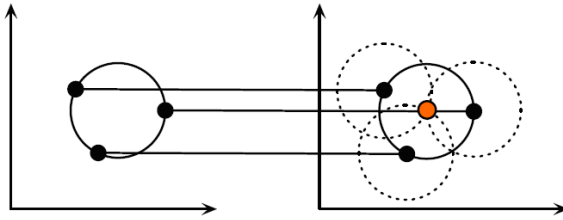


Fig. 7. Each point in geometric space (left) generates a circle in parameter space (right)

parameter circles, and can be found with a Hough accumulation array as shown in Fig. 7.

By taking advantage that both pupil and iris are round we can use Hough circle transform to find the center of pupil. Radius is in this case irrelevant.

4 OpenCV Functions Used in Our System

More detailed description of this features and functions can be seen in articles [2, 11, 6].

Object detection functions:

- `CvHaarClassifierCascade* cascade = (CvHaarClassifierCascade*)cvLoad(<classifier_filename.xml>);`
- `CvMemStorage* storage = cvCreateMemStorage([size]);`
- `CvSeq* face_rects = cvHaarDetectObjects(image, cascade, memory_storage, scale_factor, min_neighbors, flags, min_size);`

Object detection arguments:

- `image` - input image.
- `cascade` - classifier trained earlier with AdaBoost.
- `scale` - classifier cascade scale factor. typically, 1.1 or 1.2 (10% and 20%, respectively)
- `min_size` - starting minimum size of objects. By specifying large enough minimum size one can speedup processing a lot.

Image processing:

- `cvCvtColor(src, dst, [converter]);`
 - `converter` - says from which plane to which you want to convert. Ex. `CV_RGB2GRAY` means RGB to GRAY.
- `cvThreshold(src, dst, [threshold], [max_value], [threshold_type]);`
 - `threshold` - comparison value for source pixel.
 - `max_value` - source pixel can either set to 0, `src`, or `max_value` as shown in Table 1.

Table 1. cvTreshold type and operation

Threshold type	Operation
CV_THRESH_BINARY	$dis_i = (src_i > T)?M : 0$
CV_THRESH_BINARY_INV	$dis_i = (src_i > T)?0 : M$
CV_THRESH_TRUNC	$dis_i = (src_i > T)?M : src_i$
CV_THRESH_TOZERO_INV	$dis_i = (src_i > T)?M : src_i$
CV_THRESH_TOZERO	$dis_i = (src_i > T)?src_i : 0$

- cvCanny(src, dst, [canny_param_max], [canny_param_min]);
 - canny_param_min & canny_param_max - upper and lower threshold. If a pixel has a gradient larger than the upper threshold, then it is accepted as an edge pixel, if a pixel is below the lower threshold, it is rejected.
- CvSeq* circles = cvHoughCircles(gray, storage, [method], [dp], [min_dist]);
 - method - the only available now is CV_HOUGH_GRADIENT.
 - dp - resolution of the accumulator image used.
 - min_dist - minimum distance between two circles.

All of the used functions are more widely described in [7]. After successful finding of an eye ROI (Region of Interest) is set upon it so that other process filters will work only on that part of image.

```
cvSetImageROI(
    img,                /*the source image*/
    Rect(
        x,              /*rectangle start point x*/
        y,              /*rectangle start point y*/
        width,          /*rectangle width*/
        height,         /*rectangle height*/
    ));
```

5 Cursor Movement

Right after the start of the program, calibration is preformed to refer location of pupil center on mouse location. User must follow mouse cursor with eyes. It will go clockwise from the left side of the screen to the center as shown on Fig. 8. On each position there will be serious of calculations to provide average eye pupil center position while looking on the cursor.

Thanks to these 5 positions program knows width and height of visual angle. Each position is calculated for 3 sec basic on the equation:

$$x = \frac{\sum_{i=1}^i x_i}{i} ; \quad y = \frac{\sum_{i=1}^i y_i}{i} \tag{4}$$

where i is number of calculated posiions during 3 sec.

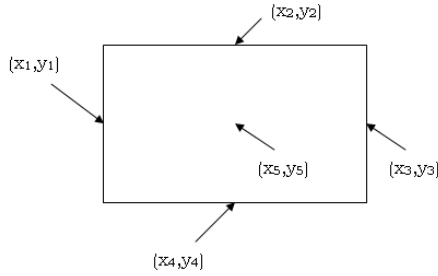


Fig. 8. Program calibration diagram

Then width and height is calculated based on:

$$width = x_3 - x_1 \quad ; \quad height = y_4 - y_2 \quad (5)$$

Fifth position is used to check calculation accretes. x_5 should merge between x_1 and x_3 , y_5 should merge between y_2 and y_4 . Basic on width and height program calculate proportion rate for mouse movement.

Although this work does not present the average number of errors per packet, the data presented in Fig. 2 shows that the period in which most errors occurred was very short, in comparison with the shortest frame duration. The most frequent value of the distance was 41 and percentage of distance less than, or equal to 96 (bit length of the shortest frame), was 83.25%. 95 percentage was achieved for distances less than, or equal to 190.

6 Conclusions

In this paper a robust method of hands free computer cursor control was presented. The system was tested against effectiveness during work on PC with MS Windows installed. During experiments we calculated both correct eye detection rate and correct pupil centre detection rate. These parameters were equal to 96 % and 93 % respectively (we used our own database of 300 images). However, it is important to note, that these parameters are calculated for single frames (images) while the system works using video-sequences. Presented errors are covered by our system while processing the consequent video frames. Therefore, the presented parameters can be only seen as technical verification of the used algorithms for single images. Now, we test our system with the help of end-users (proper validation) to receive feedback on user satisfaction, ease of use and user-friendliness. The system of eye tracking for computer control can be a good solution for physically handicapped people and when users cannot use hands. As the further work, there is plan to improve reliability by increasing camera resolution and add virtual keyboards to write documents or emails.

References

1. Al-Karaki, A.N., Kamal, A.E.: Routing techniques in wireless sensor networks. *IEEE Wireless Communication*, 6–28 (December 2004)
2. Ali-Rental, P., Ukkonen, L., Sydanheimo, L., Keskilammi, M., Kivikoski, M.: Different kinds of walls and their effect on the attenuation of radiowaves indoors. In: *Antennas and Propagation Society International Symposium*, vol. 3, pp. 1020–1023. IEEE, Los Alamitos (2003)
3. Breiman, L., Friedman, J.H., Olshen, R., Stone, C.J.: *Classification and Regression Tree* Wadsworth and Brooks/Cole Advanced Books and Software, Pacific California (1984)
4. Chipcon products from Texas Instruments, CC1101 low-power sub-1 GHz RF transceiver datasheet (2009)
5. Choraś, R.S., Kozik, R.: Head Gesture Recognition for Human Computer Interaction. In: *Image Processing and Communications Challenges*, pp. 268–275. EXIT (2009)
6. ITU-T Recommendation O.153, Basic parameters for the measurement of error performance at bit rates below the primary rate (1992)
7. Jiancheng, X., Haoshan, S., Jietai, W.: Analysis of frame length and frame error rate for lowest energy dissipation in wireless sensor networks. In: *4th International Conference on Wireless Communication, Networking and Mobile Computing, WICOM 2008*, pp. 1–4 (2008)
8. Kuranov, A., Lienhart, R., Pisarevsky, V.: An Empirical Analysis of Boosting Algorithms for Rapid Objects With an Extended Set of Haar-like Features. Intel Technical Report MRL-TR-July02-01 (2002)
9. Lodin, A., Kovacs, L., Demea, S.: Interface of an Iris Detection Program
10. Nguyen, K., Wagner, C., Koons, D., Flickner, M.: Differences in the infrared bright pupil response of human eyes. In: *Proc. of the Eye Tracking Research and Applications Symposium*, New Orleans, LA (2002)

Communications

Errors Nature in Indoors Low Power 433 MHz Wireless Network

Bartosz Boryna¹, Bożydar Dubalski², Piotr Kiedrowski², and Antoni Zabłudowski²

¹ APATOR S.A, Toruń,
e-mail: bartosz.boryna@grupa-apator.com.pl

² Institute of Telecommunications,
University of Technology & Life Sciences,
Kaliskiego 7, 85-796 Bydgoszcz, Poland
e-mail: Bozydar.Dubalski@utp.edu.pl,
Piotr.Kiedrowski@utp.edu.pl,
Antoni.Zabludowski@utp.edu.pl

Summary. In this paper, we present some results of long-term measurement in the 433 MHz, wireless telemetric network dedicated to watt-hour meter reading. The RSSI of received packets was measured by the sink. The sink was equipped with a half wave, 4 dBi antenna. Received and recorded data enabled us to make analysis of PER versus RSSI from all nodes within range of the sink, simultaneously. With knowledge of how many packets sink expected to receive, during the query and how long frames were, we propose the designation of the value of FER estimator. Lack of relationship between FER and frame length prompted us to investigate errors at bits level. Both the sink and other nodes were equipped with CC1101 low power RF transceiver. The paper concludes with the areas of the results application.

1 Introduction

In Europe, electromagnetic waves with frequencies 433 MHz and 868 MHz are very often used for automatic meter reading (AMR) systems and wireless sensor networks (WSNs). Working in ISM bands with those frequencies, maximum transmission power might be only 10 dBm. There can be many kinds of things in buildings that affect error occurrence [1]. Besides, the signal weakness that is additionally attenuated by many walls and can be easily disturbed, the second reason is communication with the shadowing effect. In general, the shadowing effect makes communication possible when antennas are separated by walls and floors but it also causes errors by the signal reflection and deflection.

A test wireless network consists of 70 nodes and one sink, all located in the campus of University of Technology and Life Sciences in Bydgoszcz. The

flooding, multi-hop technique has been implemented due to limitation of the transmission range. Using flooding technique, intermediate nodes relay many of packet copies. Packets transmitted by nodes, located in the vicinity of the sink are received by it and also recorded. All the traffic of the sink vicinity was recorded in the period from October 2009 till March 2010, what gives us about 12GB of data to analyse. Recorded data contains both transmitted and received packets with errors and error free ones, together with their time stamps and received signal strength indicator (RSSI) values. The data collected can be used in many areas of research, for example for examining changes of propagation conditions.

In many works the effect of transmission errors is reduced only to describe the relationship between RSSI and packet error rate (PER) for a given length of the packet. Knowing the relationship between RSSI and bit error rate (BER), together with a distribution of errors in packets, may be very useful, e.g. in protocols designing. Obvious relationship between BER and PER (i.e. $PER=1-(1-BER)^{packet_length}$) applied in many research works e.g. [2] or [3] is a simplification, which in many cases leads to misinterpretation of the results. Known results of errors distribution in the high-power outdoor radio systems are useless in this matter. According to our knowledge, there is no work so far that describes the results of long-term out-of-service measurements in the indoors, low power, 433 MHz-based wireless communication.

In this paper the question of what the nature of errors according to $RSSI$ value is has been considered. The answer to this question allows us to find a more precise relationship between BER and PER , which is very important for designing short haul radio communication systems. The results of this work together with the results presented in [1] can also be sufficient for WSNs modelling and simulation.

This research was supported by the Polish Ministry of Science and Higher Education and APATOR S.A. Torun, Poland.

2 The Study in More Detail

As stated in the title, the network was working with frequency 433 MHz, using GFSK modulation, no Forward Error Correction (FEC) and bit interleaving were used. Transmitted frames consist of: four-byte preamble, two-byte synchronisation word and payload which carries the packet. The packet consists of length indicator byte, address field, data field and two-byte cyclic redundancy check word (CRC-16) which is calculated from bytes of length indicator and data field. The transmission bit rate was 10 kbps.

Our experimental network may be treated as a typical WSN, which nodes consists of radio module and sensor [4], which is watt-hour meter. All nodes worked in the test mode, that means that instead of real data from watt-hour meter the pseudo-random bits sequences ($PRBS$) of length 29-1 [5]

were transmitted over the packet's data field. On the basis of the fact, that the PRBS sequences are deterministic and that fields of length and address are known, it is possible to determine not only the value of PER but also the BER .

The traffic is forced and coordinated by the sink which queries nodes one by one, using twenty seven-byte packets. The destination node responds to the query, the length of response is one hundred and twenty bytes. In response, the sink transmits eleven-byte acknowledge packet. Therefore, in the network there can only occur packets of three fixed lengths which is a consequence of the fact that the network node is part of the watt-hour-meter, which has data organized with such rules. All the other nodes, which are not queried, at the moment act as a transfer nodes. Such a solution provided 98.8% level of reliability of readings under conditions of 6.44% PER , assessed by the sink.

3 Packet Error Analysis

The main aim of the long-term test was to assess reliability of the reading system. It was also interesting the PER character, both as a function of RSSI and packet length. PER , was calculated using the formula as below:

$$PER = \frac{e_p}{ef_p + e_p} \quad (1)$$

where: ef_p is the number of received error free packets and e_p is the number of received erroneous packets.

During the whole period of the test, sink received packets with power from -115 dBm up to -17 dBm. This range can be divided into three subs: from -115 dBm up to -110,5 dBm where errors always occurred, from -110 dBm

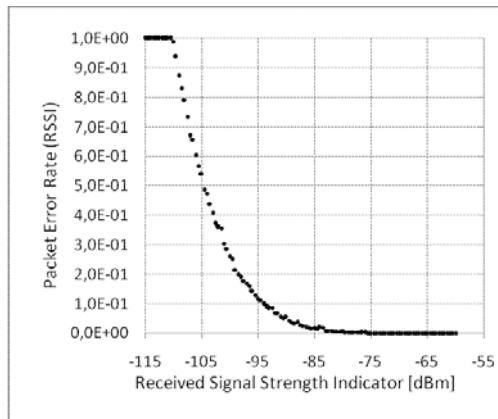


Fig. 1. PER versus RSSI measured at the sink

up to -60,5 dBm where errors sometimes occurred and from 60 dBm up to -17 dBm where errors never occurred.

Results of data analysis in shorter periods were always the same, it is because there are always some nodes close to the sink and some far from the sink, regardless of the conditions of propagation.

We examine *PER* values as a function of the *RSSI* separately for a three lengths of packets i.e. 11-byte, 27-byte and 120 byte. As a result of the examination, we obtained three curves with shapes which we expected. On the other hand, we did not expect that they were not shifted in relation to each other along the *RSSI* values. Since the curves overlap, Fig. 1 only shows the curve, which presents *PER* values as a function of the *RSSI* for all packet lengths.

Obtained results prompted us to further study the impact of packet length on the value of *PER*. We calculated three values of *PER*. For the shortest packets, *PER* was about 5.53% for medium length packets, *PER* was about 6.96% and for the longest packets, *PER* was about 8.23%. *PER* values are the greater the larger the packet length, which is normal. However, we should also consider why the differences of *PERs* are so small. We should also remember that packets are carried over frames. *PER* is the ratio of erroneous packets and all the received ones, but there were also some packets which were not received at all because of errors in the frame overhead. It led us to make an analysis on the frame layer.

4 Frame Error Analysis

It is impossible to use the relation of $PER=1-(1-BER)^{packet_length}$ for estimating a new value for longer or shorter data blocks, because the *BER* describes just the condition of a link, and in our experiment the *PER* was determined for all the links in the vicinity of the sink. *FER* value can be calculated from:

$$FER = \frac{e_f}{ef_f + e_f} \quad (2)$$

Where ef_f is the number of received error free frames and e_f is the number of received erroneous frames. If the number of not received frames is known, *FER* can also be calculated from:

$$FER = \frac{e_p + u}{ef_p + e_p + u} \quad (3)$$

where u is the number of not received frames, at all.

The data gathered by the sink contains only packets, but not frames, so we could not determine the frame error ratio (*FER*) from our experiment, but knowing the length of the frame overhead, we could estimate *FER* from the data which was used to calculate *PER*.

$$e_f = e_p + u \cong \frac{L_f}{L_p} e_p \tag{4}$$

where L_f is the length of the frame and L_p is the length of the packet. Thus, an estimator of FER is described by formula:

$$\widehat{FER} = \frac{\frac{L_f}{L_p} e_p}{e f_p + \frac{L_f}{L_p} e_p} \tag{5}$$

Using (5), we calculated the three values of \widehat{FER} . For the shortest packets it was about 8.29%, for medium length packets it was about 8.38% and for the longest packets \widehat{FER} about 8.61%.

5 Frame Error Analysis

For further analysis, the data set was reduced by eliminating those packets in which the error or errors occurred in the first byte i.e. byte specifying the packet length. Having such set, it is possible to determine a number of parameters, for instance, such as: minimum, maximum or average numbers of bit errors per packets, or the distance to the next error. However, in order to explain the reason why the FER or PER is independent of packet length, we are interested in the distance between the first and the last occurrence of an error in the packet. Therefore, the data set was further reduced by eliminating those packets in which the error occurred only once or never.

The percentage of packets (with the whole, reduced set) versus the distance between the first and the last occurrence of an error is shown in Fig 2.

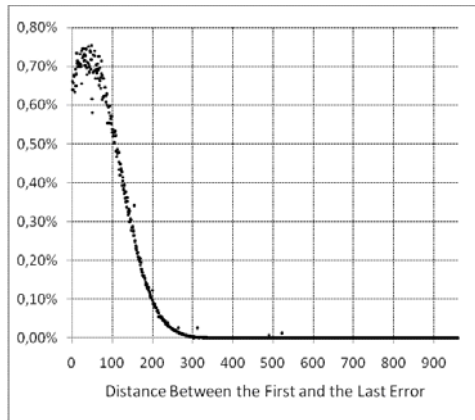


Fig. 2. Percentage of packets versus the distance between the first and the last error occurrence

Although this work does not present the average number of errors per packet, the data presented in Fig. 2 shows that the period in which most errors occurred was very short, in comparison with the shortest frame duration. The most frequent value of the distance was 41 and percentage of distance less than, or equal to 96 (bit length of the shortest frame), was 83.25%. 95 percentage was achieved for distances less than, or equal to 190.

6 Conclusions

As a result of long-term experiment, we obtained the dependence of *PER* versus RSSI in indoors conditions, which can be useful in many areas of wireless application, particularly in assessing the reliability. We expected such results, but we did not expect that the value of *PER* is practically independent of the packet length. Thus, a direct conclusion, and in fact the recommendation is such that it is better to transmit one longer packet instead of a few shorter ones, using a low power indoors radio systems. No dependence on the packet length is probably due to the fact that the share of external interference is much smaller than its own, occurring during the packet transmission through different kinds of walls, which reflect and deflect a weak signal.

References

1. Ali-Rental, P., Ukkonen, L., Sydanheimo, L., Keskilammi, M., Kivikoski, M.: Different kinds of walls and their effect on the attenuation of radiowaves indoors. In: International Symposium Antennas and Propagation Society, vol. 3, pp. 1020–1023. IEEE, Los Alamitos (2003)
2. Jiancheng, X., Haoshan, S., Jietai, W.: Analysis of frame length and frame error rate for lowest energy dissipation in wireless sensor networks. In: 4th International Conference on Wireless Communication, Networking and Mobile Computing WICOM 2008, pp. 1–4 (2008)
3. Chipcon products from Texas Instruments, CC1101 low-power sub-1 GHz RF transceiver datasheet (2009)
4. Al-Karaki, A.N., Kamal, A.E.: Routing techniques in wireless sensor networks: a survey. *IEEE Wireless Communication*, 6–28 (December 2004)
5. ITU-T Recommendation O.153. Basic parameters for the measurement of error performance at bit rates below the primary rate (1992)

Using Google Earth for Visualization in FTTH Network Planning

Jens Myrup Pedersen, Gustav Helgi Haraldsson, and M. Tahir Riaz

Aalborg University, Department of Electronic Systems,
Fredrik Bajers Vej 7, DK-9220 Aalborg East,

e-mail: jens@es.aau.dk,

gustav.haraldsson@gmail.com,

tahir@es.aau.dk

Summary. In this paper we show how Google Earth can be used for visualizing network plans, in particular for FTTH networks, supporting automated GIS-based planning by reducing the need for physically visiting sites in order to inspect and verify the plans. We develop and present a simple procedure for converting Danish GIS data into the KML/KMZ formats viewable by Google Earth, and show that it produces mappings which are precise enough for the current planning purposes. With the conversion tool in place, Google Earth is easy to use, free of charge and a strong tool for presenting and reviewing network plans. Since the same formats can be used also for Google Maps and Google Maps Mobile, it is possible to view the data from handheld devices, providing a tool that is available on-site during the network deployment.

1 Introduction

Geographical Information Systems (GIS) tools have proved their worth in making planning of larger networks more automatic, and thus faster, cheaper and more efficient than the traditional manual planning methods. In Denmark, they have been extensively used for the planning of Fiber To The Home (FTTH) networks, which are a huge infrastructure investment [1, 2]: In order to cover an area an amount of digging similar to the length of the road networks is required. The tools are used for storing, analyzing and displaying geographical and spatial information together with relevant attributes. For FTTH networks this means that network plans can be digitally represented in GIS data, which shows where the equipment is to be placed, where the cables should be dug down, etc.

The strengths however, have been mainly in the abilities to store and analyze data. Based on a set of network requirements and assumptions (e.g. the number of users, capacity needed for each user, longest distance between amplifiers, and penetration rate) together with background GIS data (roads as potential paths, the location of different kinds of users), it has been possible to create algorithms which optimize the network planning [1, 2, 3]:

The optimal topology can be calculated based on assumptions on cost of the different components such as digging, equipment costs, and fiber costs, and it can be calculated how many fibers are needed in the different road segments. Thus, automatic planning of large areas has been made possible.

However, the automated planning is not perfect. First of all, the GIS data are not always fully exact and precise. Therefore it is necessary to pre-process and clean the input data, and once the planning is completed it must be checked that the locations of ducts and equipment are feasible. This can be done using different maps and aerial photos of the areas to be planned, or by physical inspection of selected locations.

Applications such as Google Earth, Google Maps and Google Street View [4] have become popular tools for many different kinds of applications, e.g. NASA Earth Science Data [5], insurance data [6] and even for studying animal behaviors [7]. The launch of these services have also increased the possibilities for doing inspections without the need to manually fetch aerial photos or having to physically visit the locations: By presenting the network plans in a format so it can be viewed using these freely available applications from Google, it is possible to check the planning more carefully from the office. Presenting the plans in this format also makes it easier for local customer representatives or interest groups to view and comment the plans, and the mobile viewing capabilities can support the works done on the sites when the networks are deployed.

In this paper we present a simple way of translating network plans into a format viewable by Google Earth as well as the other mentioned Google applications. The paper is organized as follows: In Section 2 we recall how network planning can be done using GIS data, and summarizes the most important features of data storage and processing. Section 3 introduces Google maps and associated data formats. In Section 4 we show how the network planning data is translated into a format that can be understood by the Google applications. Section 5 concludes the paper and gives an outlook to where this can be used now and in the future.

2 The Network Planning Process

This chapter reviews the network planning process, while a more thorough description can be found in e.g. [8], and a more in-depth description of the FTTH technologies can be found in e.g. [9]. Fig. 1 shows an overview of the planning process. As a first step, the original data are collected and registered, forming the original database. It varies from country to country where these data are available, in Denmark they can generally be obtained from the local municipalities. They should be as complete as possible, and contain information about road segments, paths, railroads and other potential ducts and barriers for digging. They also contain information about where the potential network terminations (NTs) are located, and whether the NTs are business or private users. In short the objects in a GIS system are described

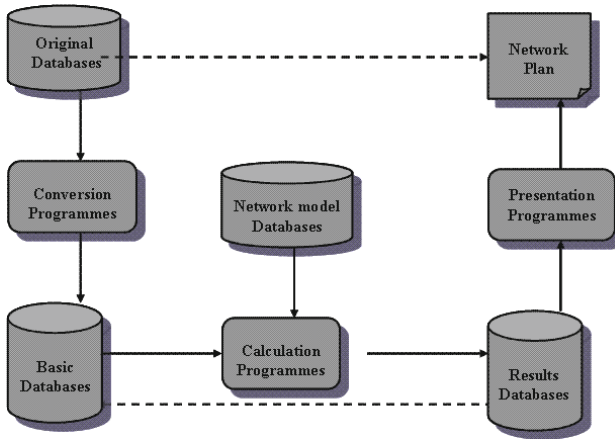


Fig. 1. Phases of GIS-based Fiber network planning

by two kinds of information: The geometry, which is the shape and location of the objects, and the attributes, which are descriptive information known about the objects. The completeness and quality of these data varies a great deal, but in general the Danish data are considered to be of high quality. At this stage the different data segments can just be stored in an appropriate database system.

In the next step, the data goes through a number of conversion programs, and scripts are carried out in order to bring the data into a standardized and common form. It is a particular challenge to automatically correct small inaccuracies such as two road segments either overlapping, or being presented as not being connected while they really are. This is very important, since the input to the algorithms later on need to be correct. At this stage, also some simplification is done to reduce the computational complexity at the later steps: Line segments are joined, and "blind roads" eliminated - in other words, the graph to be worked on is made simpler and smaller. The outcome of this process is the basic databases, which are now ready as an input to the calculation programs.

Another input to the calculation programs is the network model, which is also stored in a database. This contains information about ideal node sizes (as well as minimum and maximum sizes), maximum distances between NTs and nodes, how the paths should be chosen (e.g. shortest path or many potential customers passed), and of course also assumptions on prices of equipment as well as engineering costs.

Based on the input parameters, the calculation programs will calculate the planning parameters, mainly based on a set of graph algorithms. Once calculated, the results are stored in a result database. While the planning would be ideally "fully automatic", this is not yet the case and a number of iterations are made, each based on adjustments in the network model

and/or the design parameters. For each iteration the result is stored in a result database, and thus iterations can be carried out until a satisfactory overall plan has been generated.

At the last stage, the plan can be presented using a set of presentation programs. Usually the visual part is presented using traditional GIS viewing tools such as MapInfo or Arcview. This is then used as an input to an overall network plan, which can be presented to decision makers.

3 Google Earth and Its Data Formats

Google Earth, as well as Google Maps and Google Maps Mobile, accepts KML (Keyhole Markup Language) files, which are described and documented by Google [4] and can be found on their website. The KML file is based on the XML standard, making it fairly easy to use. As an example of a simple KML file, consider a Placemark - the well know yellow pushpin which is used to mark a point in Google Earth or Google Maps. It simply consists of a name, a description (which is shown next to the mark, often in a balloon), and a Point. The point indicates the position on the Earth, using the universal coordination system format (longitude, latitude). Thus, it is easy to insert and show points in the Google applications, and only slightly more difficult to show more advanced information such as lines, boxes etc., which can be used to show the location of the ducts/lines and equipment in a network planning context. For example, a line is characterized by a set of coordinates (which are connected by straight lines), and in the KML code a description is provided together with specifications of e.g. line length, line color, whether the line should be a straight line or follow the contours of the earth etc. An example of how a simple path is specified can be seen here:

The two most important parts of the KML code is the line style

```
<LineStyle>
  <color>ff000000</color>
  <width>5</width>
</LineStyle>
```

And the Placemark, where the line is defined:

```
<Placemark>
  <name>Simple Path</name>
  <styleUrl>#msn_ylw-pushpin</styleUrl>
  <LineString>
    <tessellate>1</tessellate>
    <coordinates>
      10.31416643322616,56.99595175503971,0
      10.31651533516604,56.99577406473386,0
    </coordinates>
  </LineString>
</Placemark>
```

Tessellate is set to 1 since we want to draw the line in multiple segments without breaks or overlaps. The resulting path is shown in Fig. 2. The example is a path along a Danish street, and it could represent the planned digging for a FTTH network. Buildings, boxes etc. can be represented using Polygons.



Fig. 2. The drawn line shown in Google Earth, as a simple example of how a network plan can be visualized

The Google applications can also work with KMZ files, which are basically zipped archives. A KMZ archive consists of a single KML file and one or more of what is called supporting files. Supporting files can contain information which is linked to the KML file. In the planning phase, this could for example contain pictures and technical information about the equipment to be installed in the different places, the amount of fiber to be dug down, or other planning related parameters. In the implementation/documentation and maintenance phases it could be used to display documentation of the work and installations.

From a Danish perspective it turns out to be a challenge though that the data comes in different formats. The majority of the GIS data available in Denmark comes in the Danish System 34 coordinate system, which is not recognized by the Google application, and no simple mathematical formula exists to do the translation. Commercially available tools such as MapInfo can do the translation, based on a transformation library acquired from the Danish Ministry of the Environment [10]. This transformation library is free to use. In the next part of the paper we explain how this conversion can be done without using the commercial tools.

4 Transformation Process

The basis for the tools is an implementation of the planning algorithms mentioned in Section 2, where the GIS data are stored in a MySQL database and the planning algorithms implemented in PHP. What is needed is therefore a process, which can be used for creating the KML files, which can then be loaded into the Google applications. The process then does the following:

- The system 34 X/Y coordinates are called from the database.
- These X/Y figures are then used as an input for the converter program.
- The converter program is executed using an external PHP system call.
- The program returns coordinates in the longitude/latitude format, and place these in the data base.
- Finally a script acquires the needed data from the database and generates a KML file based on the configuration wanted.

It is now straightforward to visualize the network plans using Google Earth as well as the other Google applications.

4.1 Application Examples

A recent FTTH planning project from the Municipality of Hals in the Northern Jutland Region of Denmark compared different network planning algorithms, leading to different network plans [11]. The first set of algorithms aimed at creating networks with simple connectivity, but no redundancy in the last mile. The second set of algorithms aimed at creating networks with more robustness, where each NT is connected to two different distribution nodes, with full redundancy at the physical level.

We used the developed transformation process to convert the final plans into KML format, in order to verify the correctness and applicability. It turned out to work satisfactory, and some screenshots from the results are provided in Fig. 3.5. For simplicity only the traces for digging and the splicing points are shown in these figures.

The results can be compared to the "traditional" presentation from Map-Info, which can be seen in Fig. 6. In the MapInfo presentation there is a perfect fit between the roads and the resulting ducts, which is not surprising since in these data sets the ducts are really a subset of the roads. The Google Earth presentation of the results is a bit more imprecise, but the meaning is still clear - both with respect to splicing and ducts - even when zooming into a specific area. For comparing the translation of the background data, we also depicted just the NTs at Google Earth (Fig. 7), and it seems to fit quite closely as well.

With an error margin of just a few meters we are not able to point to a single source for explaining it. We expect that the main contributor to the error is some slight inaccuracy the Google Earth images, but the coordinate



Fig. 3. Overview of digging and splicing points in the proposed network plan for Municipality of Hals. The map shows the western part of Hals.



Fig. 4. Overview of digging and splicing points in the proposed network plan for Hals, with a stronger zoom. It can be seen that the visualization is not fully precise, but it gives a clear picture of where ducts and splicing points are placed.



Fig. 5. Close-up of digging and splicing points in the proposed network plan for Hals

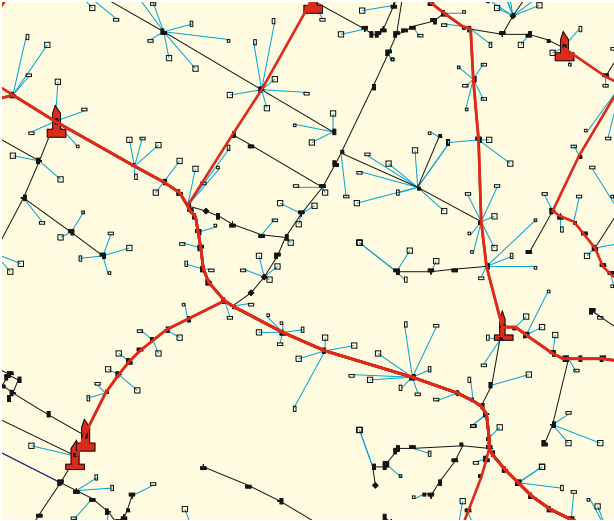


Fig. 6. Traditional presentation of a network plan (Municipality of Hals) using MapInfo



Fig. 7. The potential NTs (corresponding to houses) in the GIS database, shown in Google Maps. The mapping is quite precise. The area is in Hals.

system conversion and/or inaccuracy in the GIS data could also contribute. Google Earth does itself have a layer of streets, which seems to have a precision similar to that of the presented network plans. For the current usage, we do not see any major problems using the tools in a Danish context. However, if the planning was done in further detail (e.g. showing exactly where to dig down the cables - or even just showing which side of a road should be used) there would be a challenge.

We successfully managed to view planning data also in Google Maps and Google Maps Mobile. However, there are some rather narrow limits of file sizes for the KML files (e.g. 10 MB uncompressed), which makes it difficult to use it for larger projects. Google Street View was not used, since the streets in the Municipality of Hals are not (yet) included here. We would not expect any problems in using it though.

5 Conclusion

In this paper we have shown how Google Earth can be used as a tool for visualizing networks plans made based on GIS data. Even though the commonly used Danish GIS data format is not supported by Google it was possible to create a process that calls a conversion program, and translates the data into the longitude/latitude based KML format, which is readable by Google Earth and other Google applications.

Once this conversion has been overcome, Google Earth and the related programs Google Maps, Google Maps Mobile, and Google Street View turn out to form a toolbox with many applications in the network planning process - and they are both free of charge and easy to use. The immediate benefit is that the plans can be visualized in a tool which can be used without special instructions, and the tools can be used for checking uncertainties in the automatic planning. This is particularly beneficial in areas where the GIS data are of poor quality, so many manual corrections can be expected. It can also be used to present network plans to customers, local authorities or organizations who might suggest changes and improvements to the plans. During the implementation phase it can be accessed using e.g. Google Maps Mobile, making it accessible from the working sites. Once the networks are deployed the tools can be used to access the documentation of the networks, and KMZ files can be used for storing additional information and documentation including pictures of sites and status of the equipment. It is worth noting though that the Google applications are mainly useful for presentation, and should be seen as a supplement to existing tools.

The long term perspectives are endless. If the owners of FTTH and other types of cabled infrastructures make maps of their infrastructure available in KML/KMZ formats, it could reduce the risk for cable cuts during engineering works.

In developing countries with little documentation of paths/roads, and with no or little GIS data available, Google Earth could be used for mapping potential traces and topologies, and based on that simple scripts could support network planning and dimensioning.

References

1. Pedersen, J.M., Riaz, M.T.: Bringing fiber to the home to rural areas in Denmark. In: Proceedings of ISABEL 2009, The 2nd International Symposium on Applied Sciences in Biomedical and Communication, IEEE Conference Proceedings (2009)
2. Jensen, M.J., Nielsen, R.H., Madsen, O.B.: Comparison of Cost for Different Coverage Scenarios between Copper and Fiber Access Networks. In: Proceedings of the 8th International Conference on Advanced Communication Technology: Toward the Era of Ubiquitous Networks and Societies. IEEE, Los Alamitos (2006)
3. Poon, K.F., Mortimore, D.B., Mellis, J.: Designing Optimal FTTH and PON Networks Using New Automatic Methods. In: Proceedings of 2nd IET International Conference on Access Technologies, IET, pp. 45–48 (2006)
4. <http://www.google.com>
5. Chen, A., Leptoukh, G., Kempler, S.J., Di, L.: Visualization of NASA Earth science data in Google Earth. In: Proc. of SPIE Geoinformatics 2008 and Joint Conference on GIS and Built Environment: Geo-Simulation and Virtual GIS Environments, vol. 7143 (2008)

6. Slingsby, A., Dykes, J., Wood, J., Foote, M., Blom, M.: The Visual Exploration of Insurance Data in Google Earth. In: Proceedings of GIS Research UK 16th Annual Conference GISRUK 2008, Manchester Metropolitan University (2008)
7. Begall, S., Červený, J., Neef, J., Vojtěch, O., Burda, H.: Magnetic alignment in grazing and resting cattle and deer. Proceedings of the The National Academy of Sciences of the USA 105(36), 13451–13455 (2008)
8. Riaz, M.T., Pedersen, J.M., Madsen, O.B.: A Method for Automated Planning of FTTH Access Network Infrastructures. In: Proceedings of IT&T Annual Conference 2005, Ireland (2005)
9. Lin, C. (ed.): Broadband Optical Access Networks And Fiber-to-the-home: Systems Technologies And Deployment Strategies. Wiley, Chichester (2006)
10. Danish Ministry of the Environment, <http://www.mim.dk>
11. Haraldsson, G.H., Beshir, O.: Automatic Network Planning for Access Networks in FTTH with and without Redundancy. Master thesis, Aalborg University (2007)

The Development of a Platform Based on Wireless Sensors Network and ZigBee Protocol for the Easy Detection of the Forest Fire. A Case Study

Andreas Vlissidis, Stavros Charakopoulos, and
Emmanouil Makrygiannakis

Laboratory of Industrial Information Systems,
Technological Educational Institute of Crete, Greece

Summary. Early detection of forest fires is the primary way of minimizing their damages. Compared with the traditional techniques of forest fire detection, a wireless sensor network paradigm based on a ZigBee technique was proposed. The proposed technique is in real time, given the exigencies of forest fires. The architecture of a wireless sensor network for forest fire detection is described. The hardware circuitry of the network node is designed based on a Crossbow node. The process of data transmission is discussed in detail. Environmental parameters such as temperature and humidity in the forest region can be monitored in real time. From the information collected by the system, decisions for fire fighting or fire prevention can be made more quickly by the relevant government departments.

1 Introduction

Forest are part of the important and indispensable recourses for human survival and social development that product the balance of earth coology. However, because of some anthropogenic activities and abnormal natural conditions, forest fires occur frequently. These fires are the most serious disasters to forest resources and human environment.

In recent years the frequency of forest fires has increase considerably due to the climatic changes, human activities and other factors. In the summer of 2007, there were more than 80 people died in Greece and 670,000 acres (2,711 km²) burned because of fires. The same year in California, 500,000 acres (2,027 km²) were devastated by the flames, causing at least 14 fatalities [4]. The prevention and monitoring of forest fires has become a global in forest fire concern in forest fire prevention organizations. Currently forest fire prevention methods largely consist of patrols observation from watch towers and lately satellite monitoring. The first method its easy and feasible has several defects. First requires many financial and materials. The use of a satellite detection

system is also restricted by a number of factors which reduces its effectiveness in forest fire detection. For example cloud layers may mask images during the scanning period. This noise not allows the qualification of fire data parameters to achieve. Another problem is the long time scanning of the area and the resolution of the image pixels are very low. Moreover, satellites usually cannot forecast forest fires before the fire is spread uncontrollable [1]. Consideration the above for the data monitoring we decide to develop a monitoring platform, using the Wireless.

2 Sensors Networks (WSN) Technology and the ZigBee Communication Protocol

In recent studies, Wireless Sensor Networks (WSN) has been proposed for fire detection [2]. In WSN research community, selection of sensors was often carried out randomly or assumption-basely. Although temperature sensors are probably the simplest and the most obvious sensors for fire detection, studying various sources in this field reveals that all researchers agree on the fact that it alone is not a suitable indicator for fires and gas concentration sensors result in a better fire detection and discriminating fire and noise sources discriminating fire and noise sources [3].

In our approach, we adapt the optimal sensor set, use temperature, humidity, CO CO₂ etc, sensors. The information is transferred wireless through a 2.4 GHz channel. One of the biggest advantages of this solution is the cost which is small. So this will be the main technology that we are going to use to build our system. We assume that every node in the WSN contains all the required sensors. In this case, communication overhead between neighbouring nodes is avoided and each sensor node can detect fire locally by itself. To achieve this goal, sensor nodes need a computationally cheap, yet, efficient algorithm to conduct fire detection in a (near) real-time manner.

3 The Platform

We can point out 3 main parts in the system:

- the Communications Network and Protocols
- the Wireless Sensor Network
- the Reception Center

3.1 The Communication Network

Each node will be equipped with a receiver and a transceiver. We had to choose the protocol of communication we were going to use. After research on the available protocols for this kind of communication we decided to use the ZigBee protocol. ZigBee is a universal protocol (it was developed by 300 companies so their products can communicate) for devices to speak and interact with each other. Some of the features of ZigBee are:

- Easy to install and maintain.
- Expandable. New nodes can entered the network and interact with the existing nodes easily
- The networks size can vary from tenths to thousands of nodes
- Open source protocol. There is a big community that evolves the protocol and fixes bugs or adds new features
- The usage of bandwidth is being done with a way that ensures long life battery usage on the node
- Use the ZigBee Mesh technology for the nodes communication

Other Protocols Available

Other available protocols for data transfer are: Bluetooth and 802.11 a/b/g/n where these protocols focus on large data transfers in small amount of time, but they also use a lot of power to achieve that. WiFi protocol uses IP protocol for addressing the devices so there is also the problem of max devices in a single network. Someone can make multiple networks to communicate but this needs extra devices to be added in the network (routers) that will be taking the role of routing data between networks. This makes the network difficult to maintain and expand. Bluetooth on the other has developed for use between mobile devices mainly. Max distance between devices is 10-20 meters (Bluetooth v1.3) and one of the biggest disadvantages is that each node can only "speak" with one node at a time.

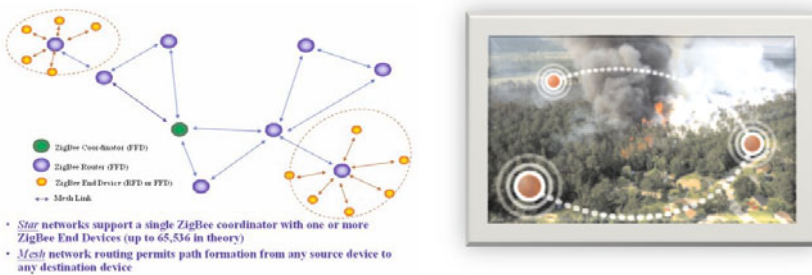


Fig. 1. The topology of a ZigBee Network and Application on fire detection

The Laboratory of Industrial Information Systems has experience of WSN networks and WSN nodes of previous projects that involved data acquisition and processing data from nodes. So these are the specifications of the nodes that we are going to use [2, 5].

Each node will be consisted from two parts. The CPU board and the sensor board The heart of CPU board will be an ATMEL ATmega 128L. On the board will be also a communication chip that will be responsible for the communication role of the node. The chip will be a IEEE 802.15.4 compliant

Table 1. The topology of a ZigBee Network and Application on fire detection

Here is a table comparing the discussed protocol			
Market Name Standard	ZigBee 802.15.4	Wi-Fi 802.11b	Bluetooth 802.15.1
Application Focus	Monitoring & Control	Web, Email, Video	Cable Replacement
System Resources	4KB-32KB	1MB+	250KB+
Battery Life(days)	100 - 1000+	5-5	1-7
Network Size	Unlimited	32	7
Bandwidth (KB/s)	20 - 250	11000+	720
Transmission Range (meters)	1 - 100+	1-100+	1-10+
Success Metrics	Reliability, Power, Cost	Speed, Flexibility	Cost, Convenience

RF transceiver working on the bandwidth 2.4 - 2.48 GHz with a data rate up to 250kbps. The board will be equipped with a 51 pin connector where the sensor board will be connected to.

3.2 The Wireless Sensor Network

A Wireless Sensor Network (WSN for short) is a distributed system consisting of a base station and a number of wireless sensors nodes endowed with radio transceivers.

Wireless Sensors networks belong to the class of Ad-hoc networks but they have specific characteristics that are not present in general Ad-hoc networks. WSN have severe energy, computation, storage and bandwidth constrains. A popular low -end node design from U Berkeley , the MicaII uses a 7.37 Mhz 8-bit Atmel CPU , with 128 Kb of FLASH Memory, only 48 KB of RAM, and a 32 Kbps Chipcon radio. Tme major resource problem in such a network is energy because these are static unattended networks and the nodes cannot have renewable energy sources. Energy is so important that algorithms designed for WSN often sacrifice response latency, accuracy, and other user desired qualities to save energy and prolong operational life time of the network.

The Laboratory of Industrial Information Systems has experience of WSN networks technologies and design and develop applications using WSN to collect and process environmental data and presenting them in real time on the Internet. So these are the specifications of the nodes that we are going to use [\[5\]](#).

The development of our fire detection system is based in the Micaz nodes of the firma Crossbow.

Each node will be consisted from two parts. The CPU board and the sensor board

- The heart of CPU board is the Atmel ATmega 128L. On the board will be also a communication chip that will be responsible for the communication role of the node. The chip will be a IEEE 802.15.4 compliant RF transceiver working on the bandwidth 2.4 - 2.48 GHz with a data rate up to 250kbps. The board will be equipped with a 51 pin connector where the sensor board will be connected to Fig. 2

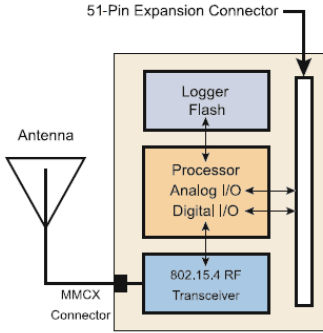


Fig. 2. The block diagram of the Micaz node



Fig. 3. The Sensor Board

- For our application, on the sensor board selected the following sensors:
 - For temperature the Intersema MS5534AM with a range of -10 to +60 Celsius degrees of measurements.
 - For humidity we will be using the Sensirion SHT11 sensor that can give us a range from 0% to 100% humidity with a 0.1% step accuracy
 - For lighting measurement we will be using the TAOS TSL2550D
 - and last for measuring the CO we will be using the GMM 221 sensor. The main chemical derivative of fire is CO and that is the reason is using such a sensor.

3.3 The Reception Center

The reception center is equipped with a central node and a USB to PC connection. The central node (MIB520) receives this message and if everything is ok with the data forwards the packet to the computer through the USB2 SERIAL port which it is equipped. There an application has the duty to here if new data arrives in the specified port a grab them to post process them. This application is called Xserve. It is an application written in C language.

4 Programming the Nodes

The nodes described above support the TinyOS operating system. TinyOS is an open-source operating system designed for wireless embedded sensor

networks. It features a component-based architecture which enables rapid innovation and implementation while minimizing code size as required by the severe memory constraints inherent in sensor networks. So what we get with the TinyOS is a platform where we can program using a high level programming language like NesC (a C based language with object oriented features). We don't have to care about memory management, or how the message is going to be delivered, these issues where taken care from the operating system. One advantage of the TinyOS is that we can simulate code we have written using one of the many simulators that have been developed. An example we have in the Fig. 4. Another advantage is that it is not an operating system that depends on a language, so we can use any operating system we want to develop our software and test it.

```

dark /root/src/tinyos-2.x/apps/Blink -f-> make telosb
mkdir -p build/telosb
compiling BlinkAppC to a telosb binary
ncc -o build/telosb/main.exe -Os -O -mdisable-hwmmul -Wall -Wshadow
-DDEF_TOS_AM_GROUP=0x7d -Wnesc-all -target=telosb -fnesc-ofile=build/telosb/app.c
-board= _BlinkAppC.nc -lm
  compiled BlinkAppC to build/telosb/main.exe
      2782 bytes in ROM
       61 bytes in RAM
msp430-objcopy --output-target=ihex build/telosb/main.exe build/telosb/main.ihex
writing TOS image

```

Fig. 4. TinyOS source code compiling

```

configuration OscilloscopeAppC
{
}
implementation
{
  components OscilloscopeC, MainC, ActiveMessageC, LedsC,
    new TimerMilliC(), new DemoSensorC() as Sensor,
    new AMSenderC(AM_OSCILLOSCOPE), new AMReceiverC(AM_OSCILLOSCOPE);

  OscilloscopeC.Boot -> MainC;
  OscilloscopeC.RadioControl -> ActiveMessageC;
  OscilloscopeC.AMSend -> AMSenderC;
  OscilloscopeC.Receive -> AMReceiverC;
  OscilloscopeC.Timer -> TimerMilliC;
  OscilloscopeC.Read -> Sensor;
  OscilloscopeC.Leds -> LedsC;
}

```

Fig. 5. TinyOS. NesC source code.

5 System Architecture

As mentioned before, the use of wireless sensor network technology in real-time forest fire detection. Our goal is to detect and predict forest fire promptly and accurately in order to minimize the loss of forests, wild animals, and people in the forest fire. In our proposed paradigm, a large number of sensor nodes are densely deployed in a forest. Sensor nodes collect measured data (e.g., temperature, relative humidity, CO...) and send to their respective cluster nodes that collaboratively process the data by constructing a neural network. These

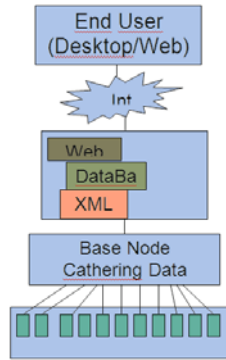


Fig. 6. The block diagram of system architecture

sensor nodes are organized into clusters so that each node has a corresponding cluster header. Every sensor node sends measurement data, as well as the location information, to the corresponding cluster head, using for example a GIS system or a GPS system. The neural network takes the measured data as input to produce *weather index*, which measures the likelihood for the weather to cause a fire. Cluster headers will send weather indexes to a manager node via the sink. Then the manager node concludes the forest fire danger rate based on received weather indexes and some other factors. In certain emergent situations, sensor nodes may inform the reception center (Manager Node) Fig. 6. A few wind sensor nodes are manually deployed over the forest and connected to the sink via wired networks to detect wind speed.

The manager node provides two types of information to users: (1) emergence report for abnormal event (e.g. smoke or extremely high temperature is detected); (2) real-time forest fire danger rate for each cluster based on the weather indexes from the cluster header and other forest fire factors. Users can also query the current temperature and humidity data in particular cluster area.

Because the system is focusing on nodes of WSN, the data and the conclusions of the data should be very specific and precise. One issue that we have to be very careful is the problem of the nodes' energy consumption. Each node should consume the lowest amount of energy so the batteries on the node live longer. Energy is being consumed when a node is trying to send a data packet to another node and when a node is trying to process the measurements in real time.

For the first we decided the time intervals to be long enough to ensure that no critical environmental change that takes place in the monitored area is lost. Also when a node "catches" a environmental variable that is out of the normal values, starts to monitor this specific value more intensively. This feature is also used as an alarming system for our system. The main server when a node starts this procedure is being informed and starts a sequence of actions such as emailing or sms the system administrator and system operators about the

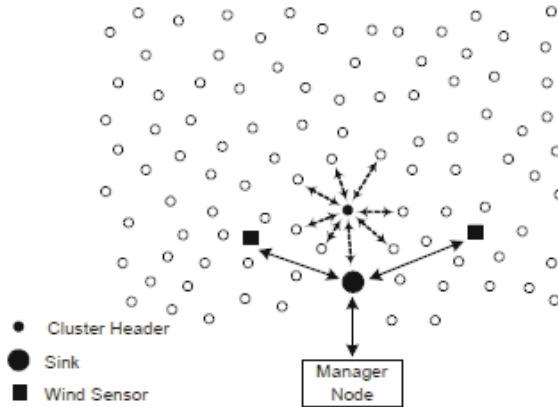


Fig. 7. A wireless sensor network for real-time forest fire detection

situation, with included information like where the event took place , how long this is happening and what type of alert we have. For reducing the cpu resources that a node will need to process the data that has acquired we are going to use the CPU power of the server to process these data.

These two simple methods reduce the amount of power used on one node. Measurements of how much this reduction is are currently being made and we don't have valid data until now.

6 Presentation of Data

Each time a nodes send data to the main server these data are being store in a data base. From there we can present these data to the end users (Web and desktop). The users can extract historical data from a simple interface in



Fig. 8. A image with on line presentation of one node

from of data indexes or data graphs. In our system we have integrated a GIS system that will assist the system operators to have a better view of where an event is taking place. The GIS system we used is the Google map API and each node has been placed on the map on its real coordinates. Fig. 8. The system can be communicate with other systems too because is the messaging system is based on XML services.

References

1. Liyang, Yu., et al.: Real-time Forest Fire Detection with Wireless Sensor Networks. 0-7803-9335-X/05/\$20.00 IEE (2005)
2. Tan, W., et al.: Mine Fire Detection System Based on Wireless Sensor Network. In: International Conference Information Acquisition, ICIA 2007 (2007)
3. Milke, J.A.: Using Multiple Sensors for Discriminating Fire Detection. In: Fire Suppression and Detection Research Application Symposium (1999)
4. <http://www.libelium.com/>
5. On line data and graphs from the WSN platform, Bioclimatic House TEI Crete/Heraklion, <http://bioklima5.prv3.teiher.gr/crossbow/live.php>

Useful sites

<http://www.springerlink.com/content/167347334n258772/fulltext.pdf>

<http://www.zigbee.org/LearnMore/Tutorials.aspx>

<http://docs.tinyos.net/index.php/>

<http://www.xbow.com/>

Mazovia Broadband Network (MBN Network). Case Study

Antoni Zabłudowski, Bożydar Dubalski, and Łukasz Zabłudowski

Instytut Telekomunikacji UTP Bydgoszcz

e-mail: {antoni.zabludowski,dualski,zabludowski}@utp.edu.pl

1 Introduction

Currently, the broadband networks, co-financed with the European funds, are being deployed in many different regions of Poland. In previous financing period, that took place in years 2004 - 2006, the only one broadband network, namely in Kujawsko - Pomorskie region, has been successfully implemented. In the current financing period the broadband networks are intended to be implemented in some other regions of Poland. The main goal of the regional broadband networks implementation in Poland is to fulfill the gap which appeared as a lack of investment in broadband networks infrastructure.

OECD statistics [1] for 3rd quarter of year 2009 shows that among 30 examine world countries Poland is on 28th place regarding population of broadband networks. The average population of broadband access links in Poland are at level 11 ones per 100 citizens where 7 of them are implemented in ADSL technology and the other 4 are implemented in DOCSIS technology (cable TV). In Polish network there are almost no FITL technology access links, what is a result of limited investment in fiber infrastructure in all layers of the network.

One of the regions of Poland that is on the stage of preparing broadband network conception is Mazovia region situated in the central part of Poland with Warsaw as the capital of region. The Mazovia network conception assume to build 1500 kilometers of passive optical cabling, to install of active equipment (IP routers, ME switches) in backbone and distribution layer as well as to implement of NGA solutions in certain parts of the region. The presented paper is focused on technical conception of MBN network backbone layer.

Broadband Internet is the basis for the development of all modern networks and first of all consists a base for the construction of NGN networks. The strength of the Internet aptly characterized Giovanni E. Corazzi, Professor, University of Bologna, who in his conference presentation on 19 September

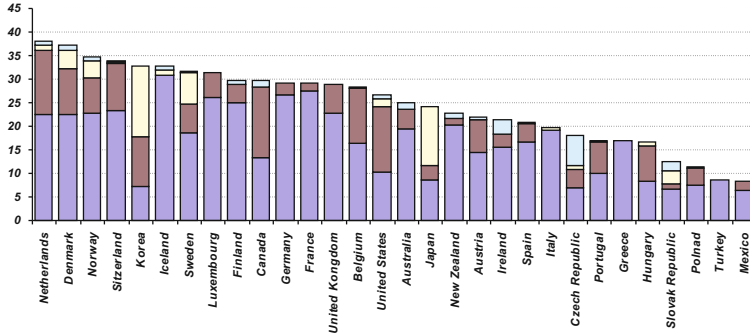


Fig. 1. OECD Broadband subscribers per 100 inhabitants, by technology, June 2009

2008, at Unic Workshop in London [2], put several fundamental theses concerning the Internet, of which four major are listed below:

1. in 2012 the Internet will be 75 times bigger (in terms of total volume traffic) than the Internet in 2002;
2. is the privileged who have access to the Internet;
3. Internet provides easy penetration of information between different social groups deployed around the world;
4. up to 2020 entire world knowledge will be located and reachable on the Internet.

In 2012 acceptable access link bandwidth for Mazovia voivodeship regional broadband network (used as the essential assumption for network planning) should be from 6 to 10 Mb/s, but in 2015 and later from 20 up to 100 Mb/s. The access links in MBN should be symmetrical with the very similar bandwidth for both directions. High-speed symmetrical access links, requires to use the optical technology for "last mile before" part of access network, but in the near future the optical access lines should be applied in "last mile" part also.

2 Estimation of Total Volume IP Traffic Carried in MBN Network

The total volume of traffic carried by the network is determined by the usage of applications implemented on MBN network. Services offered in network are deciding both about the network layers realization and the income gained from network operation. Networks based on IP protocol realize services including: WWW, Peer to Peer, streaming (i.e. IP telephony, IP TV, VoD), networks chats (text and voice) [3]. New types of IP networks services directly impacts on network architecture.

Cisco reports shows [4, 5] that the biggest growth of traffic volume is followed by residential users that are offered the following services:

- symmetrical access to global Internet with bandwidth not less than 15 Mbit/s;
- peer to Peer services (P2P);
- IP telephony, video telephony and internet radio;
- IP TV, at the beginning SDTV, latter HDTV and in the future including 3D standard;
- video on demand (VoD);
- work based on home office with broadband access;
- e - government services.

For local government, business and security the following services are implemented:

- 112 emergency number;
- other (except 112 number) shared number for undefined situations, similar to 311 number in United States or 115 in Germany;
- monitoring of selected region zones and schools;
- integration of emergency in case of extreme situations including assigned VPN's for rescue services for communication;
- e - learning;
- e - health.

Services offered in Internet network generates traffic which has to be carried in specific logical layers of the network. The volume of traffic is different for different kind of services so the total amount of traffic distribution is essential to be known. At this moment the biggest volume of traffic is generated by the P2P services and in the future by video services including video for computers and for TV's. The Cisco reports [4, 5] says that three elements of IP traffic which is P2P, total video traffic and classic internet traffic (WWW, HTML, e - mail) is slightly more than 96% of traffic volume generated by residential users in the network. P2P traffic is decreasing its percentage share in total traffic volume from about 60% (year 2006) to slightly about 33% (year 2012), but in a range of traffic volume it grows 5 times. The similar situation is for WWW traffic. It decreases its volume in total Internet volume traffic from 22,3% (year 2006) to level 15,1% (year 2012), despite this the total volume of traffic increase 6 times. Considerably increase traffic volume generated by video to TV receivers. Comparing year 2006 and 2012 it shoots up 250 times [4, 5, 6].

For estimating efficiency of active equipment installed in designed network it is essential to know where the specific traffic regarding specific services flows in the network. This especially refers to quantification if the network traffic is either shut in the scope of designed network or directed to IXP's (Internet eXchange Point). The network examination shows that 80% of total traffic carried in regional network is directed to IXP points. It refers the Peer

to Peer traffic, WWW traffic, video traffic for PC's and traffic generated by video games. The element of global IP traffic which is IP television (i.e. IP TV or VoD) is generally shut among regional network and only slight amount of this traffic is directed to other IP networks (which are IXP's). Virtually whole traffic related to e - health services and security services is transferred among regional network.

Traffic generated by specific class of services has a fundamental impact on access links - either symmetrical or asymmetrical. Services like IP TV, VoD, video for PC, and WWW generates asymmetric traffic which volume is directed to the bottom of the network that is to end users. This traffic is much bigger than traffic volume directed to the top of the network (IXP's or service servers). Likewise traffic generated by monitoring services is an asymmetric traffic. Services like P2P, VPN networks, tele - presence VoIP, services of communication for security management in region, e - government or e - health services generate symmetric traffic.

The most remarkably and most growing element of whole traffic volume is a traffic generated by residential users. That means that when defining requirements of MBN network the requirements of residential users have to be considered. That is why in assumption of MBN network plans the links for residential users are all symmetric. The traffic generated by business users, local government or in example e - health services is lower than traffic generated by residential users but it still in most cases symmetric. This is why MBN network assume that all business users will be linked by symmetric connections. As MBN network is still to be build, planning and offering symmetric connections are simplify.

For estimating traffic volume generated in MBN network whole region of Mazovia has been divided for separate distribution zones where backbone node and some distribution nodes are installed. The arrange of distribution zone has been shown in Fig. 2. The enumeration of traffic generated in each distribution zone has been realized under following assumptions: the average bandwidth of access link is 15 Mb/s, the average population of MBN network users in distribution zone is 25%. The yellow points presented on the map of Mazovia region shows the location of distribution points where users are linked in by direct connections using either star topology or using intermediate module (hierarchical star topology). Aggregated traffic from distribution nodes flows to backbone layer nodes next it hits the core of network.

Concerning carried assumptions regarding MBN network users the total traffic generated in the network is about 85 Gb/s. Taking under consideration the Cisco report about traffic growth in networks [4, 5] we can assume that in the next 10 years the traffic in the network grows to 3,7 Tb/s. Such quantity of traffic is need be carried out by efficient equipment especially in backbone and core network layer. Whole network traffic generated in MBN network is directed to three core nodes that gives 1/3 (about 1,2 Tb/s) for each, however the destination size of network traffic is so big that efficiency of MBN network routers is a unconditionally must.

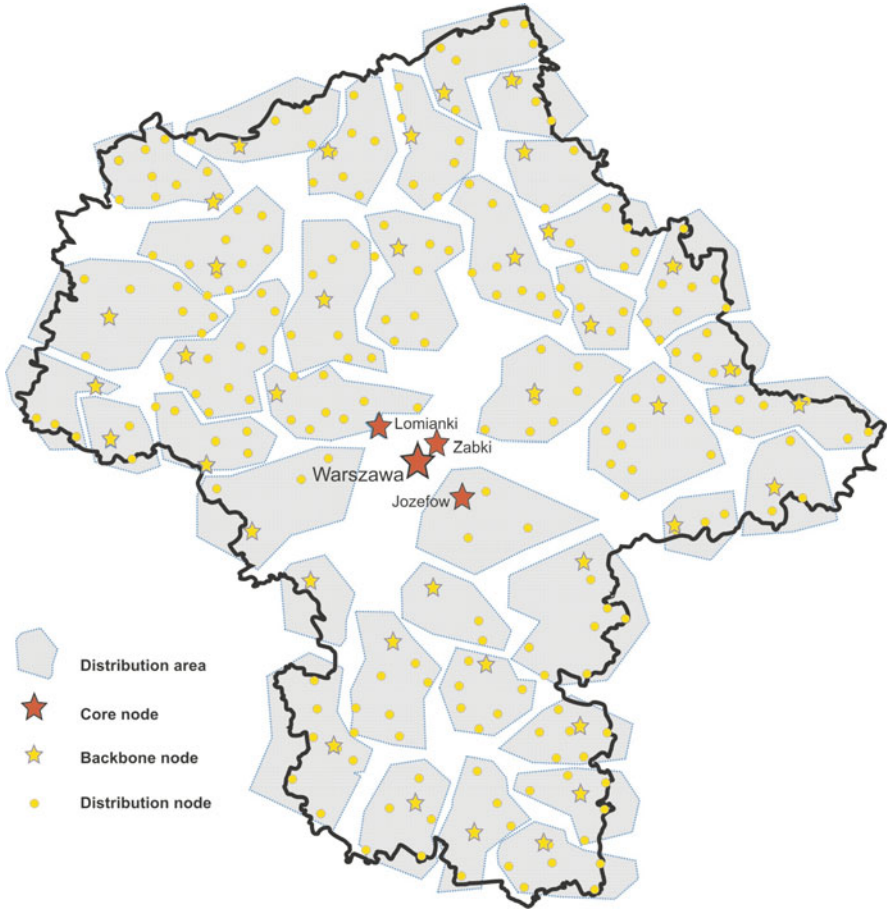


Fig. 2. Distribution zones of MBN network

3 MBN Network Implementation

The topology structure of the passive fiber physical layer for MBN network has been presented in Fig. 3. This figure shows built fibers including backbone and distribution layers. From available fibers in MBN network the best topology structure has been created. In this case three sections which goes through Warsaw are using two pair of fibers (4 fibers) which has been gain under IRU (Indefeasible Right of Use) role from commercial ISP's which are present in Mazovia region. This fibers sections on IRU license are: link Warsaw - Jozefów, link Warsaw - Łominaki, Warsaw - Zabki.

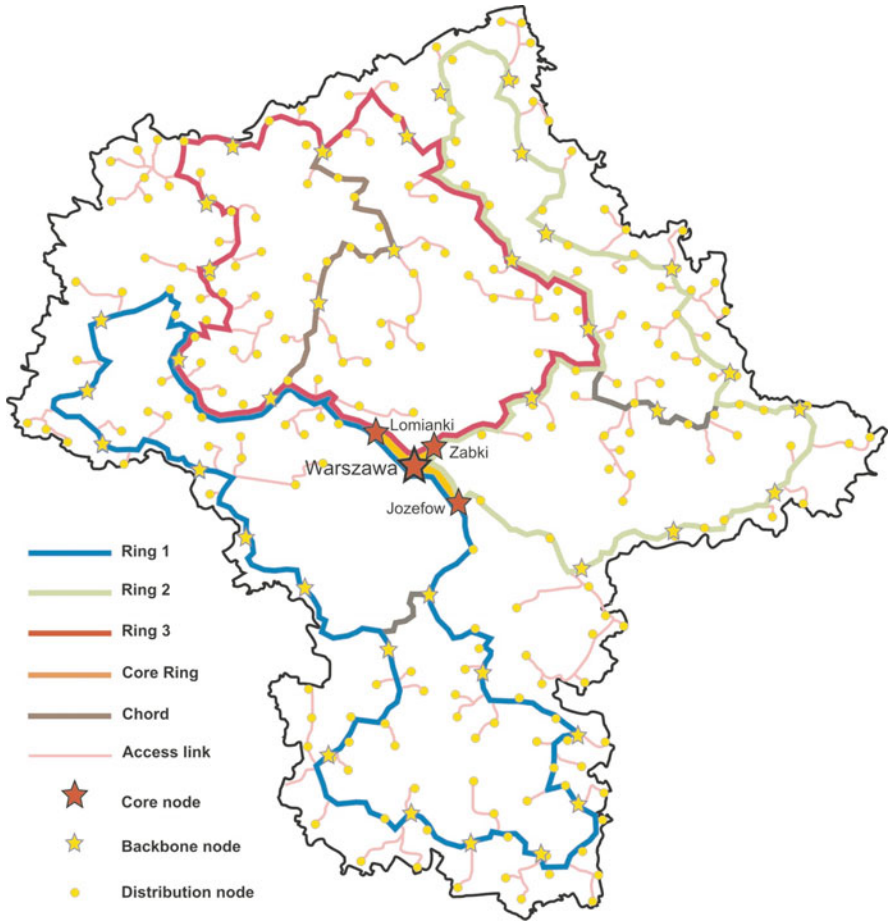


Fig. 3. Topology structure of passive fibers layer

Fig. 3 reveals that passive backbone network is a Mesh structure, where transfer equipment is installed in backbone nodes. The best and the most safe topology of fiber network is the ring structure. The ring structure brings easy reconfiguration of the network when equipment failure appears. This is why even Mesh networks are implemented on the mechanism of ring structure defining special sub networks of Mesh networks called p - cycles. These p - cycles are used to predefining and using, in a case of an equipment or link failure, the security paths for traffic transfer in this cycle. The MBN network passive structure topology has been built as three rings structure with one chord in each ring. This solution has been applied not only because of the specification of broadband network traffic transfer but also because in such networks the fibers redundancy occurs. Because three rings topology was

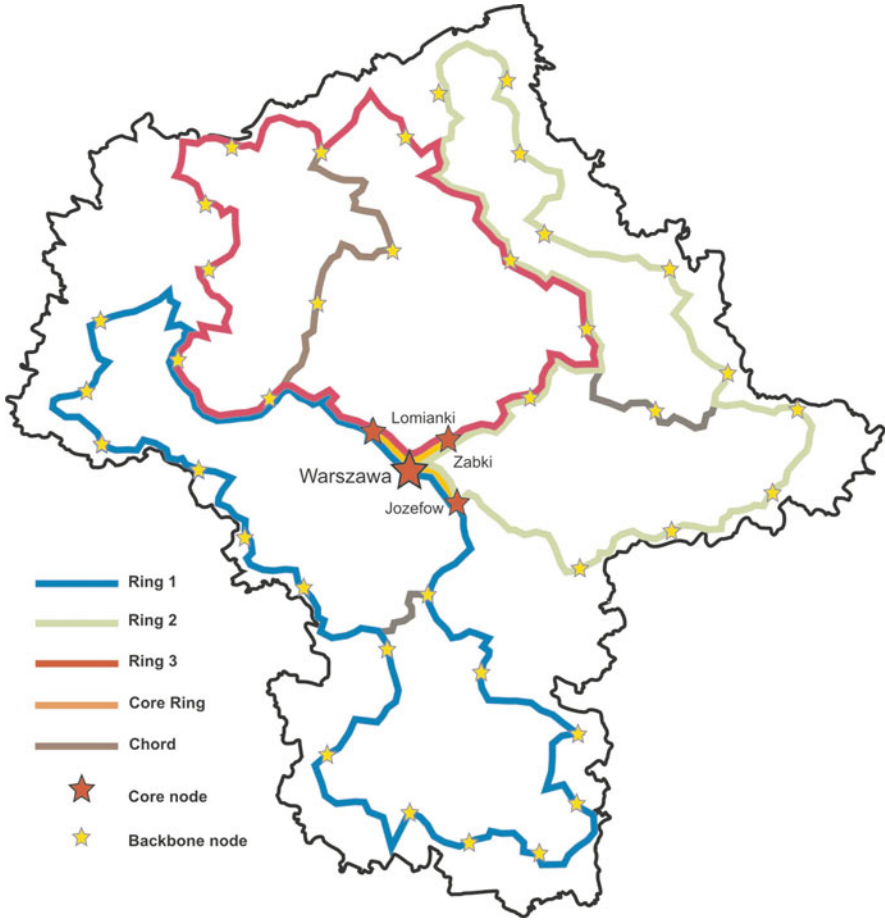


Fig. 4. Three rings as a topology of physical MBN backbone network layer

used when designing MBN network the p - cycles mechanisms are being used already at designing stage.

In our feeling the most optimum topology implementation for MBN network is a structure of the network presented in Fig. 4. Each ring of MBN network is a double fiber ring so 1 + 1 protection can be easily implemented. To avoid unnecessary increase of network cost the channels placed in the chords of the rings are not 1 + 1 protected but for path protection in the chords there is a possibility of traffic transfer in both directions.

In three ring MBN network the shared path problem has to be resolved. The shared path exists at a joint of two rings and is implemented on four fibers, two fibers for each ring. This path involves some backbone nodes where

two rings go through them. Those nodes suppose to be connected with core nodes. For linking shared nodes one ring is enough. The only problem is to connect those nodes to core nodes layer. There is no reason to install two pieces of active equipment (i.e. routers or OADM) in each access node existing on the shared path because for linking distribution node to core layer nodes only one active piece of equipment is needed. However for the second ring there is a necessity to estimate the link budget and place optical amplifiers if needed. When using optical DWDM systems existing of a shared path for two rings allows to optimize the number of optical channels in the rings.

Aggregation core nodes which main role is to aggregate the traffic from each ring and transfer it to central Warsaw node have been placed in nodes of network which exists in two rings and are connected directly to central Warsaw node. MBN network summary core nodes has been placed in: Jozefów, Łominaki and Ząbki.

To provide high reliability of network core layer, all core nodes (aggregation nodes and central Warsaw node) are connected into a ring using Mesh structure in a way that this complete Mesh structure consists of two aggregation nodes and one central Warsaw node. The optical channels which link aggregation core nodes and central Warsaw node in a ring are $1 + 1$ protected, which means that a complete three nodes network exists. That gives three complete Mesh structures where each aggregation core node consolidates to two such structures. In a ring the one following node exists: Jozefów and Łomianki, in the second ring: Jozefów and Ząbki and in the third ring: Łomianki and Ząbki. Because each ring of MBN network goes through Warsaw central node therefore profiting with potentiality of DWDM optical transport system the logical structures of the network has been possible to build where each two aggregation core nodes and central Warsaw node form three node Mesh network.

4 MBN Architecture and Traffic Transfer in the Network

The traffic transfer in MBN network precisely depends on architecture of this network. This is why before the basis of the MBN network traffic is presented the general architecture of modern broadband IP networks is shortly going to be discussed which has been implemented in MBN network. The architecture of broadband IP network consists of:

1. Core layer which is a central part of every telecommunication network and is delivering services for users connected to this network.
2. Backbone layer which is connecting meshed network routers and realizes transfer between core layer and users located in the access layer.
3. Access layer which role is to connect, using backbone layer, users to core layer. In access layer we can distinguish at least two functional sub layers

which are: distribution sub layer (aggregation) - last mile before and subscriber sub layer - last mile.

United Europe regulations determine that EU grants regarding regional broadband networks can be used only to build all three layers of network that is: core layer, backbone layer and access layer however for this last layer only in a range of distribution sub layer.

Presented architecture of broadband IP network in the following essay determines only functionality of the network seen by the prism of traffic transfer. For the goal of the project the construction of the core layer is not important. All you need to know is the location of service servers, the sort of services in this network and volume of traffic generated for each of service. The core layer of the broadband IP network acts very important and special role as its implemented on selected network nodes where has been installed access points to different types of services. Generally the regional broadband network core layer offers two classes of services:

- Services offered by service servers (i.e. VoD, IPTV, etc.)
- Services of access to global Internet provided by IXP points.

The most characteristic thing for IP networks is that almost 100% of traffic is being directed to core nodes where 80% of traffic is directed to IXP points. Such traffic transfer in IP networks which fundamentally is directed to core network enforces special structure (physical or logical) of backbone network which is called by us as a structure physically oriented. The structure's main goal is to support traffic transfer to network nodes where traffic is transferred. In the case of broadband IP network where the majority of traffic is directed to core nodes the structure should be build in the way that the shortest path exists between backbone nodes and core nodes. With experience in the construction of K-PSI leads to the following conclusion regarding the construction of backbone and core layer:

- the best logical topology of the backbone layer of the broadband IP network physically oriented is a star structure connecting the core nodes with backbone nodes. This structure can be easily implemented using xDWDM equipment.
- the topology of the core layer should be designed as a Mesh structure (complete structure as best).

5 MBN Network Options and Active Equipment Configuration

For designing MBN network needs three conceptions of active equipment installation has been elaborated. In conclusion presented in the previous sections consider general assumptions have been defined at the stage of defining the active network layer architecture.

1. MBN network is build on the three optical rings structure whereby each of the rings is at average of about 13 backbone nodes, where two of which remain core functions of aggregation nodes.
2. The backbone nodes aggregate the traffic from distribution nodes and transfer it to network core.
3. Aggregation core nodes intermediate the IXP's and provide access to service servers installed in central Warsaw node. In aggregation core nodes are installed MPLS routers where half the traffic generated in the optical ring is directed.

Considering the functionality and efficiency of the MBN individual nodes the following types of nodes, depending on the conceptions, has been set:

1. Central node - **router type A_1** which efficiency is not less than 20 Tb/s which is connected to three aggregation core nodes using 40 GE interfaces. The central node does not serve the access layer.
2. Aggregation core nodes - **routers type A_2** which efficiency is not less than 5 Tb/s, installed in three MBN network points. Depending on option routers are equipped with either three or one 40 GE interface. Interfaces facing access layer are 1 GE each.
3. Nodes placed in the rings chords - **routers type A_3** which efficiency is not less than 500 Gb/s. Depending on conception, routers are equipped with either 40 Gb/s or 10 Gb/s interfaces. Interfaces facing access layer are 1 GE each.
4. Routers or Ethernet switches acting as a MBN backbone nodes - **modules type B_1** . Their role is to transfer traffic from access nodes to core nodes. Efficiency of B_1 type equipment is not less than 500 Gb/s. Interfaces facing access layer are 1 GE each.
5. Ethernet switches which are **modules type C_1** are used for connecting access nodes and to aggregate traffic from this nodes and transfer it to distribution nodes.

Option 1 of MBN Network

- The network is build on the IP MPLS routers platform equipped with 40 Gb/s interfaces.
- In backbone and core layer the following equipment is installed: one router type A_1 (central node), three routers type A_2 (aggregation core nodes), five routers type A_3 (nodes in the chords of the rings) and thirty three routers type B_1 (backbone routers). All routers are equipped with 40 Gb/s interfaces and support IP MPLS protocol.
- C_1 nodes are installed in network distribution layer, that are MetroEthernet switches supporting IP MPLS protocol.
- IP MPLS routers are installed in ring's backbone nodes and network traffic is transferred by all routers situated on the route from distribution node (backbone) to aggregation core node.



Fig. 5. Nodes connection in MBN network according to option 1

- In each aggregation core node (type A₂) IXP performing equipment is installed.
- In one of the ring of the shared path between two rings the optical amplifiers are installed.

The MBN network architecture build accordance with option 1 has been showed in Fig. 5.

Option 2 and 3 of MBN Network

- The network on the physical level of transport layer is build with a use of DWDM with 32 optical channels. The optical equipment is installed

only in core layer and backbone layer and they do not appear in access layer. For logical network layer the OADM modules are used. The optical channels provide information transfer at level of 10 Gb/s.

- In core layer and backbone layer of the network the following equipment is installed: router type A_1 (central node) equipped with 40 Gb/s interfaces, three routers type A_2 (aggregation core nodes) equipped with 40 Gb/s interfaces for central node connection purposes and 10 Gb/s interfaces for communication with backbone nodes, five routers type A_3 (nodes in rings chords) with 10 Gb/s interfaces (option 2) or five MetroEthernet switches (option 3). All routers and MetroEthernet switches installed in option 2 and 3 support IP MPLS protocol.

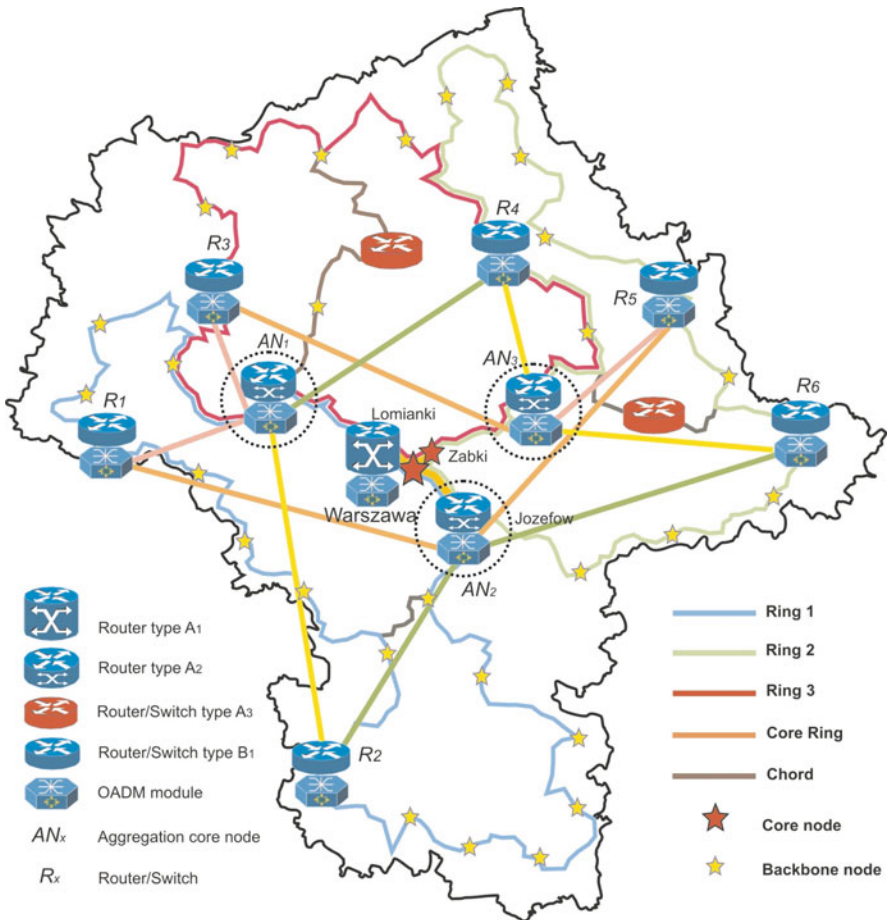


Fig. 6. Nodes connection in MBN network according to option 2 and 3

- IP MPLS routers are installed in core and backbone nodes equipped with 10 Gb/s interfaces which provide scalability of the network for at least 10 years. Optical channels of DWDM system form logical structure of the star where every node of backbone layer situated in one of the three rings is connected using one dedicated optical channel with central node with bandwidth 40 Gb/s.
- Traffic from backbone nodes is transferred to core nodes using direct optical channels.
- In distribution layer of option 1 case nodes type C_1 are installed which are MetroEthernet switches supporting IP MPLS protocol.
- IXP performing equipment is installed in every core node type A_2
- In one of the ring of the shared path at the interface of two rings the optical amplifiers are installed.

The MBN architecture build accordance with the option 2 and 3 has been shown in Fig. 6. This figure also explains the principle of connection implementation in the network for option 2 and 3. In the network node, each of the rings, the single optical OADM module and IP MPLS router (MetroEthernet switch) is installed. Network users are linked to the router using distribution nodes of different hierarchy. The aggregation core node (type A_2) is situated in two rings and backbone nodes of option 2 solution are linked to two aggregation core nodes using optical channels with a bandwidth of 10 Gb/s.

6 Construction Cost for the Various Network Options

The comparison of the different kind of network realization has been made. For all options the access layer (closely the distribution sub layer of this layer) is the same and build using ME switches. Also central core node is using the same configuration for all three options. The MBN network options differ from each other in respect of network backbone realization and comparison of the cost refers to nodes A_2 , A_3 and B_1 . The Equipment price and specification is represented in Tab. 1. Each cost of MBN network realization option contains the cost of purchasing the managing system.

Option 1. Routers type A_2 are equipped with three 40 GE interfaces, which are used for connecting routers type A_2 with a router type A_1 and neighboring routers type A_3 , and four 10 GE interfaces for connecting nodes type C_1 . In the option 1 network five routers type A_3 are installed (routers in the chords of the rings) equipped with three 40 GE ports (towards routers type A_2 and A_3) and four 10 GE ports for connections with C_1 nodes and thirty routers type B_1 (routers in ring's backbone nodes) equipped with two 40 GE ports (towards routers type A_2 , A_3 and B_1) and four 10 GE ports for connecting routers type C_1 .

Option 2 and 3. Routers type A_2 are equipped with wit one 40 GE interface which is used for connections of routers type A_2 with A_1 and twenty eight 10 GE interfaces for connections with nodes type C_1 installed in every

ring's node. Nodes type \mathbf{A}_2 are equipped with 28 10 GE ports for connections with nodes type \mathbf{A}_3 and \mathbf{B}_1 . In the MBN network there are installed five routers type \mathbf{A}_3 (routers installed in the ring's chords) which are equipped with six 10 GE ports and thirty routers type \mathbf{B}_1 (routers in the ring's network backbone nodes) which are equipped with two 10 GE ports (for routers type \mathbf{A}_2) and twenty eight ports 1 GE ports towards access switches type \mathbf{C}_1 . In both options 36 DWDM (OADM modules) pieces of equipment are installed. These modules work using optical channels with 1 + 1 protection which are connecting nodes type \mathbf{A}_2 with nodes type \mathbf{B}_1 . In option 3 the MBN network is build exactly in the same way as network in option 2 but instead of nodes type \mathbf{A}_3 and \mathbf{B}_1 which represents IP/MPLS routers the MetroEthernet switches has been used. This configuration of the network does not change much in connections of the network but it force different type of equipment logical configuration.

Option 2 and 3. Routers type \mathbf{A}_2 are equipped with wit one 40 GE interface which is used for connections of routers type \mathbf{A}_2 with \mathbf{A}_1 and twenty eight 10 GE interfaces for connections with nodes type \mathbf{C}_1 installed in every ring's node. Nodes type \mathbf{A}_2 are equipped with 28 10 GE ports for

Table 1. Equipment price and specification for network building options

Option 1										
Node type	Quantity type	Equipment	Ports				Price List [USD]			Total
			40G	10GE	GE	10/100	Equipment	Cards	Optics	
A ₁	1	Router	3				117 000\$	2 250 000\$	0\$	2 367 000\$
A ₂	3	Router	3		20		351 000\$	2 345 000\$	597 000\$	3 293 000\$
A ₃	5	Router	3		20		585 000\$	2 345 000\$	597 000\$	3 527 000\$
B ₁	33	Router	2		20		3 861 000\$	1 650 000\$	656 700\$	6 167 700\$
C ₁	270	Switch			2	24	3 508 650\$		537 300\$	4 045 950\$
	1	NMS/OSS					1 946 670\$			1 946 670\$
							10 369 320\$	8 590 000\$	2 388 000\$	21 347 320\$
Option 2										
A ₁	1	Router	3				117 000\$	2 250 000\$		2 367 000\$
A ₂	3	Router	1	28			351 000\$	1 275 000\$	336 000\$	1 962 000\$
A ₃	5	Router		6			585 000\$	190 000\$	144 000\$	919 000\$
B ₁	33	Switch		2	28		990 000\$		264 000\$	1 154 000\$
	35	DWDM					17 691 070\$			17 691 070\$
C ₁	270	Switch			2	24	3 508 650\$		537 300\$	4 045 950\$
	1	NMS/OSS					2 900 140\$			2 900 140\$
							26 142 860\$	3 715 000\$	1 281 300\$	31 139 160\$
Option 3										
A ₁	1	Router	3				117 000\$	2 250 000\$		2 367 000\$
A ₂	3	Router	1	28			351 000\$	2 850 000\$		3 201 000\$
A ₃	5	Switch		2	28		150 000\$		48 000\$	198 000\$
B ₁	33	Switch		2	28		990 000\$		264 000\$	1 154 000\$
	35	DWDM					17 691 070\$			17 691 070\$
C ₁	270	Switch			2	24	3 508 650\$		537 300\$	4 045 950\$
	1	NMS/OSS					2 944 400\$			2 944 400\$
							25 752 120\$	5 100 000\$	849 300\$	31 701 420\$
Internet Exchange Point										
IXP	3	Router		12			351 000\$	900 000\$		1 251 000\$

connections with nodes type \mathbf{A}_3 and \mathbf{B}_1 . In the MBN network there are installed five routers type \mathbf{A}_3 (routers installed in the ring's chords) which are equipped with six 10 GE ports and thirty routers type \mathbf{B}_1 (routers in the ring's network backbone nodes) which are equipped with two 10 GE ports (for routers type \mathbf{A}_2) and twenty eight ports 1 GE ports towards access switches type \mathbf{C}_1 . In both options 36 DWDM (OADM modules) pieces of equipment are installed. These modules work using optical channels with 1 + 1 protection which are connecting nodes type \mathbf{A}_2 with nodes type \mathbf{B}_1 . In option 3 the MBN network is build exactly in the same way as network in option 2 but instead of nodes type \mathbf{A}_3 and \mathbf{B}_1 which represents IP/MPLS routers the MetroEthernet switches has been used. This configuration of the network does not change much in connections of the network but it force different type of equipment logical configuration.

7 Conclusion

The analysis of network options realization shows that the most low cost of MBN network building solution is option 1. This solution uses only routers in a backbone layer which transfer traffic from aggregation nodes to core nodes. This solution however decreases the network performance because all routers carry transit traffic. Despite that optical interfaces of this routers are scaled at 40 GE it may be that in the next couple of years the interfaces bandwidth are not efficient enough and their replacement with 100 GE interfaces is necessary. In options 2 and 3 this weakness does not appear because DWDM system allows to build physically oriented networks where traffic is directed to core nodes. DWDM system also allows to build logical structure linking backbone nodes with core nodes using direct optical channels. Each backbone router transfers to network core only aggregated traffic (coming from access layer) in this router. If in a ring exists 11 routers of this kind which are linked using direct channels of 10 Gb/s with two core nodes than for handling traffic generated by such network, in option 1, the routers with 100 GE interfaces should be installed. Despite the realization cost in option 2 and 3 is higher by over 50% the option 2 for implementing MBN network is recommended as a solution that meets the broadband network expectations.

References

1. OECD Broadband statistics. OECD Broadband subscribers per 100 inhabitants, by technology (2009)
2. Corazza, G.E.: University of Bologna. Who can bridge the information divide? In: Unic Workshop, London (September 19, 2008)
3. Vanston, L.K., Hodges, L.R.: Technology forecasting for telecommunications. Telektronikk, TELENOR 4 (2004)

4. Cisco: Cisco Visual Networking Index: Forecast and Methodology, 2007-2012 (June 2008)
5. Cisco: Cisco Visual Networking Index: Forecast and Methodology, 2008-2013 (June 2009)
6. Swanson, B., Gilder, G.: Estimating the Exaflood. The Impact of Video and Rich Media on the Internet. A zettabyte by 2015? Discovery Institute (January 2008)

The Method of GMPLS Network Reliability Evaluation

Janusz Korniak and Paweł Różycki

University of IT and Management in Rzeszow, Poland,
e-mail: {jkorniak,prozycki}@wsiz.rzeszow.pl

Summary. Reliability evaluation is important task in a network design process. The GMPLS network supports physical separation of functional planes and therefore reliability evaluation becomes more difficult. The method of GMPLS network reliability evaluation is proposed in this paper. The idea of this method is based on the GMPLS as a multistate system. In order to demonstrate usefulness of this approach, proposed method is implemented and reliability evaluation for test network is preformed.

1 Introduction

The GMPLS (Generalized Multiprotocol Label Switching) [4], developed by IETF, is one of the most promising technique, which is considered to use in the next-generation backbone networks. It introduces the separation of the following functional planes:

- data plane – responsible for transmission of user data; it includes all switching techniques such as WDM, TDM, packet switching, etc.;
- control plane – responsible for exchanging signaling and routing messages; it is implemented as IP network that support protocols known from MPLS such as LDP, RSVP, OSPF;
- management plane – responsible for supervise of whole system; it may be implemented as centralized or distributed system that allows to employ provider's policy.

In fact, the management plane is not a part of GMPLS standard but it seems to be required in the context of interdomain policy and multivendor environment. Moreover, GMPLS is considered as implementation of signaling mechanism in the ASON (Automatically Switched Optical Network) [6], backbone concept of ITU-T, where the management plane plays very important role.

The control plane is crucial for GMPLS network operation. The main task of the control plane is effective management of calls and connections

(establishment, deletion, maintenance) as well as management of network resources. To support these tasks, GMPLS-related protocols can be used, such as: IS-IS, OSPF, BGP for routing and CR-LDP, RSVP-TE for signaling. To manage signaling channels and data links between nodes a new protocol LMP (Link Management Protocol) has been developed.

1.1 The GMPLS Network Architecture

Separation of functional planes implies that signaling in GMPLS network may be implemented in the following manners:

- in-band – when the signaling network is a part of the data plane and signaling channels are implemented as channels logically separated from the supported transport technology (e.g., DCC within SDH/SONET);
- out-of-band – when the signaling network is physically separated from the data plane.

The out-of-band signaling allows for using different topologies in functional planes. In this case, two nodes directly connected in the data plane may not be connected in a such way in the control plane or, similarly, two nodes directly connected in the control plane may not be neighbors in the data plane. Such architecture of GMPLS is called asymmetrical. Otherwise, the architecture is called symmetrical.

The out-of-band signaling has many advantages. It avoids opto-electronic processing because user data and signaling messages are physically separated. Moreover, various schemes for protection and restoration can be applied in this case. On the other hand, failures in the data plane and the control plane have to be handled separately [1] that involves a more complicated implementation. Separation of the planes implies possibility to keep established data plane connections in the event of the control plane failure. Similarly, in the case of the data plane failure a signaling along broken connections may be still possible because the control plane is not affected by failure of the data plane network.

One of the most important issues of the GMPLS, especially when it is used as a commercial backbone network, is to ensure a high level of reliability and quality of service. The problem of GMPLS reliability, also for out-of-band signaling, is therefore, the object of many analyses.

1.2 Reliability Evaluation for Complex Systems

Reliability of network system in the context of telecommunication is defined as 'the probability that an item can perform a required function under stated conditions for given time interval' [5]. Reliability analysis is important for designing and optimization of a network. Especially for backbone network, it is desired to maximize reliability due to a very high cost of system unavailability. Therefore reliability evaluation is very important for network

engineers. There are several reliability evaluation methods. They can be classified to [5]:

- two terminal reliability – known also as source-target reliability, and
- all terminal reliability.

Form the point of view of the end-to-end services the first method is suitable because express that path between source and target is operational. There are many network reliability evaluation methods, including [7]:

- reliability block diagramming,
- state enumeration,
- monte carlo and discrete event simulation.

Reliability of optical backbone networks often is analyzed in the context of protection and restoration methods [9] or specific network architecture [3]. Discrete event simulation is often chosen for these tasks because allows scientist to include this specific assumptions in reliability evaluation. In this paper authors present universal, analytic method for reliability estimation regardless of protection and restoration methods, suitable for networks with separated functional planes as described in subsection [1.1]. That method can be very useful in initial planing of network architecture especially before mentioning of protection and restoration. Moreover, it can be faster then simulation methods.

2 GMPLS as Multistate System

Some of the reliability evaluation methods are not suitable for complex network. For example, traditional, reliability block diagramming method is popular but requires to present a network as a parallel-series system. Reduction of complex network system to parallel-series one is not effective. Because GMPLS network is obviously complex system, the effective reliability evaluation method is necessary. One of the reliability evaluation method adequate to the complex network is the overlap technique described in [2]. This method is applied to calculate source-target reliability of GMPLS network.

The main problem with reliability evaluation for GMPLS is an interaction between functional planes. Each node consists the data plane and the control plane component and there are four possible states. Table [1] presents these states. In the first case both planes are available and therefore whole node is available too. The second case is very interesting. The data plane is functioning but the control plane is unavailable. In this state established Label Switched Paths (LSP) can be preserved and the user data still can be forwarded by the data plane. However, the control plane is unavailable, thus creating a new LSP or releasing existing one is not possible. This state can be called as derated. Two last cases produce down state of GMPLS node due to data plane unavailability. As a result GMPLS node can have three states.

Table 1. State enumeration for GMPLS node or link

	case data plane	control plane	GMPLS node
1	up	up	up
2	up	down	derated
3	down	up	down
4	down	down	down

Similar situation occurs when links between nodes are mentioned. The data plane nodes are connected with the data plane links and the control plane nodes are connected with control plane links. Both, data and control plane links are physically separated. The state enumeration is the same as for the node. See Table 1. Derated state in this case means that the control plane nodes have not the control plane communication channel and the control plane tasks can not be performed while the data plane still can forward user data.

Probability P_{up} that GMPLS node or link is in the up state is:

$$P_{up} = P_{cp} * P_{dp} \quad (1)$$

where P_{cp} is reliability of the control plane component and P_{dp} is reliability of the data plane component. Probability P_{dera} that GMPLS node or link is in the derated state is:

$$P_{dera} = (1 - P_{cp}) * P_{dp} \quad (2)$$

Next formula represents probability P_{down} that GMPLS node or link is in the down state:

$$P_{down} = 1 - (P_{up} + P_{dera}) \quad (3)$$

Using three state approach, two separated layers of GMPLS network can be analyzed as one network where each component has state 'up', 'derated' or 'down'.

3 Example of Reliability Evaluation

In this section reliability evaluation for simple reference network is presented. Figure 1 depicts the analyzed network. There are two following examples of reliability evaluation form node 1 to node 9 with different topologies of the control plane.

Values of reliability are evaluated with the use of overlap algorithm described in section 2 implemented by authors in the Scilab environment [8].

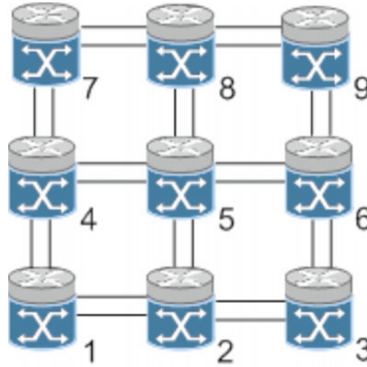


Fig. 1. The network used in examples of reliability evaluation

3.1 Example 1

In this example symmetrical topology is assumed and therefore the data and the control plane are represented by the same topology [1]. Moreover, the data plane uses some redundancy mechanisms which allow to find an alternative path in the case of primary path fail. Therefore, between node 1 and 9 any loop free path can be used. The control plane does not use any redundancy solution. Signaling between two directly connected nodes are allowed through direct link. Table 2 contains reliabilities of nodes and links assumed for the source-target reliability evaluation.

Table 2. Reliabilities of nodes and links assumed for source-target reliability evaluation

	reliability
Data plane node	0.99988
Data plane link	0.99990
Control plane node	0.99977
Control plane link	0.99990

Reliabilities of nodes and links are calculated using multi state approach described in section 2. See Table 3.

3.2 Example 2

The second example uses the same network topology as in example 1. Also the same reliabilities of nodes and links for the data and the control plane are assumed. The only difference is that the control plane supports redundancy. The signaling channel between two directly connected nodes can be created through direct link between that nodes or through any other alternative path

Table 3. Calculated reliabilities of GMPLS nodes, links and source-target

	P_{up}	P_{dera}	P_{down}
GMPLS node	0.99965	$0.22997 * 10^{-3}$	$0.12000 * 10^{-3}$
GMPLS link	0.99980	$0.99990 * 10^{-4}$	$0.10000 * 10^{-3}$
form node 1 to node 9	0.99930	$0.46030 * 10^{-3}$	$0.24010 * 10^{-3}$

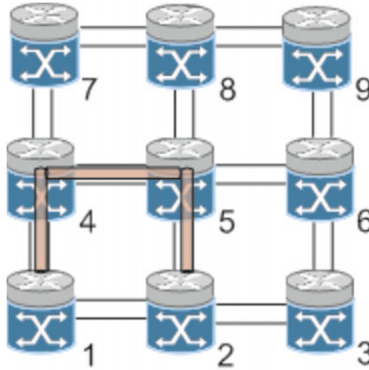


Fig. 2. Alternative signaling channel between nodes 1 and 2

between them. Example of alternative signaling channel between node 1 and 2 is shown in the Fig. 2. Other alternative signaling channel are also possible. The reliabilities of signaling channels are calculated using overlap technique and minimum value is shown in Table 4. With the redundancy, signaling channel is almost fully available. Similar to signaling channel redundancy, the control plane nodes are redundant to provide failover. Thus with redundant the control plane, node reliability is improved (see Table 4).

Using these values multi state reliabilities of GMPLS node and link are calculated. Next, obtained values are used to evaluate the source-target reliability between nodes 1 and 9. The results are summarized in Table 5.

Table 4. Reliabilities of the control plane nodes and links and calculated minimum reliability of signaling channels assuming redundancy

	reliability
Control plane node	0.99977
Control plane link	0.99990
Signaling channel	0.99999997
Redundant control plane node	0.99999995

Table 5. Calculated reliabilities of GMPLS nodes, links and source-target with signaling redundancy

	P_{up}	P_{dera}	P_{down}
GMPLS node	0.99988	$0.52894 * 10^{-7}$	$0.12000 * 10^{-3}$
GMPLS link	0.99990	$0.29997 * 10^{-8}$	$0.10000 * 10^{-3}$
form node 1 to node 9	0.99976	$0.10582 * 10^{-6}$	$0.24008 * 10^{-3}$

By comparison of last rows from Table 3 and Table 5 it is possible to conclude that redundancy in the control plane:

- improves reliability of system – probability of 'up' state is increased,
- minimizes unreliability of the control plane – probability of 'dereated' state is decreased.

4 Conclusions

The GMPLS enhances MPLS by separation functional planes. This physical separation complicates the reliability evaluation. Each component of the GMPLS network consists elements of both functional planes and user data transfer depends on both of them. However a failure in the control plane does not mean termination of user data transmission. Therefore typical methods of reliability evaluation are not suitable for this purpose.

Authors propose to use multistage approach in reliability evaluation. In this method each element can be in one of three states: 'up', 'derated' and 'down'. Derated state express the stage in which the control plane component is failed and the whole GMPLS element has limited functionality.

In order to demonstrate usefulness of this method, experimental reliability evaluation has been performed. The results show difference of reliability level for the simple network with different implementation of the control plane. Based on evaluated reliabilities it is possible to assess the impact of the control plane to the source-target (or service) reliability. Thus, this method helps in initial process design of GMPLS network.

References

1. Jajszczyk, A., Rozycki, P.: Recovery of the control plane after failures in ason/gmpls networks. *IEEE Network* 20(1), 4–10 (2006)
2. Rice, B.: Scalable complex cyber network reliability algorithm. Master's thesis, Troy University Computer Science (2009)
3. Grover, W.D., Clouqueur, M.: Availability analysis and enhanced availability design in p-cycle-based networks. *Photonic Network Communications* 10(1), 55–71 (2005)

4. Mannie, E.: Generalized Multi-Protocol Label Switching (GMPLS) Architecture. RFC 3945 (Proposed Standard) (October 2004)
5. ITU-T Recommendation E.800. Terms and definitions related to quality of service and network performance including dependability, Technical report, ITU-T (1994)
6. ITU-T Recommendation G.8080/Y. 1304 amendment 2, architecture for the automatically switched optical network (ason), Technical report, International Telecommunication Union (2005)
7. Rice, B., Sahinoglu, M.: Network reliability evaluation. *Computational Statistics* 2(2), 189–211 (2010)
8. <http://www.scilab.org/> (2010)
9. Sennhauser, U., Zhou, L., Held, M.: Connection availability analysis of span-restorable mesh networks. *Photonic Network Communications* 14(2), 135–148 (2007)

The Improved Least Interference Routing Algorithm

Ireneusz Olszewski

Institute of Telecommunications,
University of Technology & Life Sciences,
Kaliskiego 7, 85-796 Bydgoszcz, Poland
e-mail: ireneusz.olszewski@utp.edu.pl

Summary. The paper compares existing algorithms of choice of LSPs in IP networks with MPLS protocol and proposes a new algorithm, which is an improved version of the Least Interference Routing Algorithm. The study showed that in dynamic conditions of the network (only short lived connections) proposed algorithm rejects a smaller number of requests of LSPs choice in relation to the compared algorithms, irrespective of the requests rate and the network topology and size.

1 Introduction

The paper compares the algorithms given in [1, 4, 5, 6, 7, 8] and proposes a new algorithm of choice of Label Switched Path (LSP) in IP networks with the Multi-Protocol Label Switched (MPLS). The proposed Improved Least Interference Routing Algorithm (ILIOA) is an improved version of the well known Least Interference Routing Algorithm (LIOA) [1].

The considered problem concerns the choice of LSPs between a pairs of nodes in the IP/MPLS network, where the only information provided by the corresponding algorithms for LSP choice algorithm is available bandwidth on individual links in the network [2]. Algorithms based on the choice of LSP along shortest paths, which length are expressed by the number of links, are sufficient to achieve the connection, but they do not always allow full use of network resources from the point of view of traffic engineering. The main reason is that the links on the shortest paths between certain pairs of nodes: the input node - output node may be congested, while links on alternative paths remain free. This means that network resources are not sufficiently used, although it is possible to make better use of network resources on the same network infrastructure. Therefore, the search of LSP choice algorithms, which increase the use of network resources is fully justified.

Traffic Engineering is associated with the LSPs choice algorithms working both *offline* and *online*. The *offline* algorithms assume that all the LSPs,

together with the values of the bandwidth are known during the implementation of these algorithms [4]. The aim of these algorithms is the choice of LSPs in order to minimize the main criterion, which is generally the cost of the network implementation. However, in practice it is likely that new requests of LSPs set up will appear after the implementation of LSPs group or the resources required for the existing LSPs may change. The problem of the adoption of these future paths can be avoided if it is possible to redirect existing LSPs. Since the *offline* algorithms generally do not make re-allocation of demands that are already realised, and therefore when a group of LSPs is already realised and there are new requests of LSPs choice, so directing of these new paths can only be realised with the use of *online* algorithms.

The performance of *online* algorithms depends practically on information of the state of links achieved from the different protocols, such as Open Shortest Path First (OSPF) [2]. The basic idea of these protocols is that each node in the network sends information of the state of its links and its associated attributes to all other nodes. This mechanism causes that each node in the network knows the network topology. This topology can therefore be used by the individual nodes to develop their own decisions about the next step for the output node. The LSP is set up using the Resource Reservation Protocol (RSVP). Because it is difficult to obtain the delay value or information of buffer occupancy from the signalling protocol [2], and therefore, the choice of the LSP cannot depend on them. The proposed algorithm and the algorithms of LSPs choice presented in [1, 4, 5, 6, 7, 8] are based solely on state of links occupancy and the number of flows carried by them.

The remaining part of paper is organized as follows. The second part gives an overview of the compared algorithms. The third part describes the formulation of the optimization problem and proposes a heuristic algorithm solving this problem. The fourth part contains the results of simulation, obtained after the application of the presented algorithms. The last section presents the final conclusions.

2 LSP Choice Algorithms

In this paper, as in [1, 4, 5, 6, 7, 8], only the choice of LSPs with guaranteed bandwidth is considered. This does not mean that other paths metrics such as delay and packet loss cannot be considered. The best known algorithm is the min-hop algorithm [4], which chooses the shortest path measured by the number of links between the input node and the output node. It should be noted that during designation of LSP only those links with available bandwidth which are no less than the bandwidth of desired LSP are taken into account. This algorithm belongs to the Constraint Shortest Path First class, and is also named as CSPFHopCount algorithm [3].

Constraint Shortest Path First (CSPF) has been proposed as an algorithm that solves the problem of load balancing in OSPF by using the appropriate weight of links that reflect the availability of current resources of the network.

The weight of links in the algorithm is inversely proportional to the residual capacity of these links [1, 7]. However, both OSPF and CSPF poorly implement traffic engineering in a heavy load conditions [1]. Outlined algorithms based on network topology and residual capacity of links, but do not take into account the location of input and output nodes. If the LSP between a pair of nodes is chosen regardless of location of other potential pairs of nodes, the interference of this path with requests of set up LSPs that will appear in the future can occur [4]. The Minimum Interference Routing Algorithm (MIRA) takes into account the location of the other pairs of nodes (input-output) in order to minimize the amount of interference. Therefore, the choice of LSP between a given pair of nodes is realised to maximize the minimum value of the maximum flows, computed for the available bandwidth of network links, determined for other pairs of nodes [4]. An important drawback of MIRA algorithm is quite high computational complexity [4].

Another algorithm, named as the BU-MIRA, based on a minimum of interference is given in [8]. Weights of links are determined on the basis of the available bandwidth of links and the number of flows carried on these links.

In [1] the Least Interference Optimization Algorithm (LIOA) was presented. Both MIRA and LIOA minimize interference among competing flows in the network in order to determine the weights of links that will minimize the possibility of blocking LSPs. It should be noted that both of these algorithms define interference in different ways. In LIOA the interference is defined by the number of flows (LSPs) carried by a link. However, in the MIRA interference is defined as a reduction of the maximum flow between a pair of nodes after set up the LSP between another pair of nodes. In LIOA weights of links are proportional to the number of flows (the number of LSPs) realized on these links, and inversely proportional to their residual capacity. Dispersing traffic flows (LSPs) in the network and minimizing the rejected LSP requests is obtained by minimizing interference and maximizing the residual capacity of the links [1].

In Maximize Residual bandwidth and link Capacity - Minimize Total Flows (MaxRC-MinF) routing algorithm [6] the interference is determined as in LIOA. Weights of links are directly proportional to the number of flows traversing these links and inversely proportional to their residual capacity and bandwidth. This ensures avoiding links with the large number of flows carried on them (minimizing interference), choosing unloaded links and load balancing through the choice of links with large capacities [6].

In turn, in [5] the Bandwidth Constrained Routing Algorithm was proposed, in which weights of links are proportional to the usage of links and inversely proportional to the total capacity of links. So determined weights cause avoiding a high loaded (used) links and favor the choice of links with larger capacities, resulting in load balancing [5].

3 Proposed Algorithm of LSPs Choice

Let $G(N, E, B)$ be the network, where N is the set of nodes (routers) and E is the set of unidirectional links (arcs). C is m -vector of bandwidth of the links. Let n denote the number of the nodes and m the number of links in the network. Let R be a m -element vector of the residual capacity of links. Moreover, let $P_{i,j}$ be the set of all possible paths between a pair of nodes (i, j) . In addition, let $w_l(C_l, R_l, I_l)$ denoted the weight of link l . I_l is the number of flows carried on the l -th link (the size of interference). Suppose that between a pair of nodes (i, j) the request of LSP set up with bandwidth $d_{i,j}$ appears. Informally, the problem of routing optimization can be formulated as follows: Find a path $p^*, p^* \in P_{i,j}$, with a minimum weight $Wp^* = \sum_{l \in p^*} w_l(C_l, R_l, I_l)$ with a guaranteed bandwidth $d_{i,j}$, which minimizes the number of rejected requests of LSPs set up in the network.

So formulated problem is generally solved by minimizing interference of incoming request of LSP set up with requests of LSPs set up which will come in the future [4, 7, 8] or by minimizing interference of LSPs already set up in a network with coming request of LSP set up [1, 6].

The proposed algorithm is similar to LIOA, but the weights of links are defined in a more sophisticated way. Let consider the following link weight function:

$$w_l(C_l, R_l, I_l) = (1 - u_l) \frac{I_l^\beta}{C_l^{1-\beta}} + u_l \frac{I_l^\alpha}{R_l^{1-\alpha}} \quad (1)$$

where: $u_l = 1 - R_l/C_l$ is the usage of link; in turn α and β are real coefficients from range $[0,1]$.

It should be noticed that in the case of low usage of l -th link, a decisive factor of a weight of link $w_l(C_l, R_l, I_l)$ will be: $(1 - u_l) I_l^\beta / C_l^{1-\beta}$. Weight of link, with fixed bandwidth, is minimized by minimizing interference, i.e. the number of carried flows. In order to realise the LSP in the network the links with low numbers of carried flows and large capacities will be chosen. However, in the case of high usage of link the decisive factor of the weight of a link $w_l(C_l, R_l, I_l)$ will be $u_l I_l^\alpha / R_l^{1-\alpha}$ (where $I_l^\alpha / R_l^{1-\alpha}$ is weight of link in LIOA). Weight of link will be minimized by the number of flows and maximizing the residual capacity of the link. This fosters the dispersion of traffic in the network. In order to realise the LSP in the network the links with relatively low numbers of carried flows and large residual capacities will be chosen. The links load will be kept far from congestion area, so that the number of rejected requests of LSPs set up will be minimized.

The proposed algorithm of LSPs choice is as follows.

The complexity function of ILIOA

The complexity function of MIRA is $O(n^4 \sqrt{m})$ [4], whereas for remaining algorithms is $O(n^2)$. The complexity function of proposed algorithm is also $O(n^2)$.

Algorithm 5. ILIOA

Input: Network $G(N, E, B)$, vector R of residual capacities of links; vector I of flows and pair of nodes (i, j) , between which $d_{i,j}$ bandwidth units should be realized.

Output: The path between pair of nodes (i, j) with $d_{i,j}$ bandwidth units.

1. Compute the weights:

$$w_l(C_l, R_l, I_l) = (1 - u_l) \frac{I_l^\beta}{C_l^{1-\beta}} + u_l \frac{I_l^\alpha}{R_l^{1-\alpha}} \quad \forall l \in E;$$

2. Eliminate all links $(w_l(C_l, R_l, I_l) = \infty)$, which have residual bandwidth $R_l < d_{i,j}$;
3. Using Dijkstra's algorithm compute shortest path p^* between pair of nodes (i, j) .
4. Route $d_{i,j}$ bandwidth units from node i to node j along path p^* ;
5. Update vector R and vector I ; i.e. for each link $l \in p^* : R_l := R_l - d_{i,j}; I_l := I_l + 1$.

4 Obtained Results

The verification of algorithm has been made for the two networks which contain 15 and 23 nodes, respectively. The first network which topological structure is shown in Fig.1a. contains 15 nodes (routers), connected by a links with bandwidth of 12 (thin lines) and 48 (thick lines) units [1, 4]. In turn, the second one, which topological structure is shown in Fig.2b. contains 23 nodes connected by links with the same values: 12 (thin lines) and 48 (thick lines) of capacity units [1]. The capacities of links in both networks have been multiplied by an appropriate scaling factor ω . Each link is unidirectional. In this paper it was assumed, unlike in [4] that each node can be input and output node. Therefore, in the first network 210 pairs of nodes can be distinguished, while in the second one 506 pairs of nodes.

The simulation results obtained on the basis of the presented algorithm ILIOA were compared with the results obtained on the basis of six other algorithms: Constraint Shortest Path First (CSPF), Minimum Interference Routing Algorithm (MIRA), Least Interference Optimization Algorithm (LIOA) Maximize Residual bandwidth and link Capacity - Minimize Total Flows (MaxRC-MinF), Bandwidth Constrained Routing Algorithm (BCRA) and BU-MIRA.

Network simulation was made using Monte Carlo method. It was assumed that the stream of requests of LSPs set up between each pair of nodes (i, j) is Poissonian with intensity λ and holding time of the LSP is exponentially distributed with the mean value $1/\mu = 1$. Bandwidth of LSPs is uniform distributed from 1 to 4 units. The network simulation has been done in static and dynamic conditions. In static condition LSPs are only set up, but in dynamic condition LSPs are set up and disconnected. In static conditions for given trial of the simulation, all algorithms are verified for the same stream of requests LSPs choice. In dynamic conditions the results are recorded after obtaining a equilibrium state of system. Both in static and dynamic

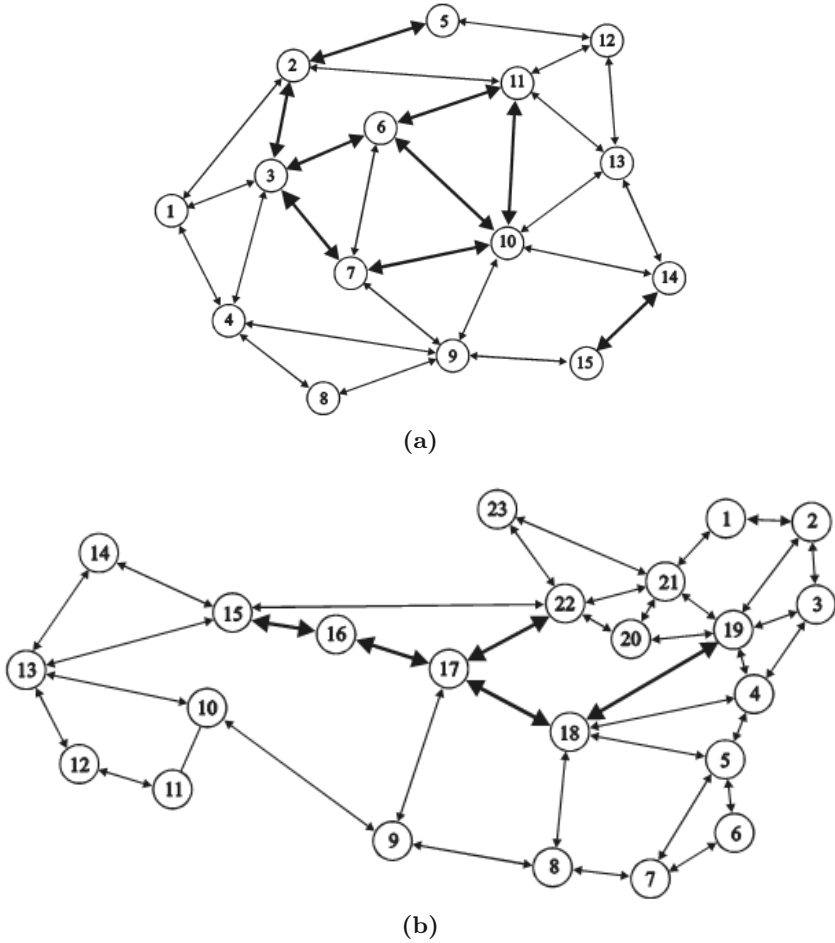


Fig. 1. Network topology: (a) the MIRA network (b) the LIOA network

conditions tests have been done for T trials. The number of requests of LSP set up in the network is a condition for the end of the simulation.

Table 1 shows the results obtained after using all the algorithms for networks containing 15 nodes, whereas Table 2 presents results for a network containing 23 nodes. In both networks the same coefficient for scaling capacity was used: for the network operating in static conditions (long lived connections) $\omega = 100$, whereas for networks operating in dynamic conditions $\omega = 10$.

During the simulation of any network, for each algorithm, $T = 20$ trials has been done. In the dynamic conditions for a network consisting of 23 nodes (LIOA network) each trial included 50000 requests, while for the network containing 15 nodes (MIRA network) each trial included 25000 requests. In

Table 1. Number of rejected LSPs for network consisting of $n = 15$ nodes

Long lived connections						
Algorithm	$\lambda=5, T=20, \text{Length Run} = 12000$			$\lambda=6, T=20, \text{Length Run} = 12000$		
	Average number of Rejected LSPs	Standard deviation	% rejected LSPs	Average number of Rejected LSPs	Standard deviation	% rejected LSPs
MIRA	339	65	3.3	431	64	3.5
BU_MIRA	232	61	0.9	165	38	1.3
CSPF	93	47	0.7	112	47	0.9
BCRA	171	27	0.6	108	50	0.9
MaxRC-F	97	49	0.8	114	50	0.9
LIOA	88	46	0.7	108	50	0.9
ILIOA	90	47	0.7	110	49	0.9
Short lived connections						
Algorithm	$\lambda=5, T=20, \text{Length Run} = 25000$			$\lambda=6, T=20, \text{Length Run} = 25000$		
	Average number of Rejected LSPs	Standard deviation	% rejected LSPs	Average number of Rejected LSPs	Standard deviation	% rejected LSPs
MIRA	969	114	3.8	2468	106	9.8
BU_MIRA	232	61	0.9	1619	113	6.4
CSPF	226	57	0.9	1589	125	6.3
BCRA	171	27	0.6	1483	150	5.9
MaxRC-F	211	67	0.8	1624	130	6.4
LIOA	161	54	0.6	1441	129	5.7
ILIOA	151	42	0.6	1425	105	5.7

turn, in static conditions each trial included 12000 requests regardless of the size of the network. Moreover, it was assumed that in the LIOA coefficient $\alpha = 0.5$ [1], whereas in the ILIOA these coefficients are $\alpha = 0.5, \beta = 0.3$.

From Table 1 it can be shown that for *long lived* connections LIOA and ILIOA reject almost the same number of requests for each value of the intensity of requests ($\lambda = 5, \lambda = 6$). In turn, in the case of *short lived* connections ILIOA rejects the smallest number of requests for each load (intensity). LIOA rejects slightly larger number of requests. Similar dependences were obtained for a network consisting of 23 nodes (Table 2). In the case of the static performance of the network MaxRC-MinF, LIOA and ILIOA are equivalent for the same intensities of requests. However, in dynamic conditions ILIOA is better than LIOA and MaxRC-MinF and the other examined algorithms.

It should be noticed that ILIOA, LIOA and MaxRC-MinF, equivalent in the case of *long lived* connections (Table 2), are based on the interference defined in the same way.

Table 2. Number of rejected LSPs for network consisting of $n = 23$ nodes

Long lived connections						
Algorithm	$\lambda=3, T=20, \text{Length Run} = 12000$			$\lambda=4, T=20, \text{Length Run} = 12000$		
	Average number of Rejected LSPs	Standard deviation	% rejected LSPs	Average number of Rejected LSPs	Standard deviation	% rejected LSPs
MIRA	1714	69	14.2	1680	60	14.0
BU_MIRA	1857	61	15.4	1836	74	15.3
CSPF	1420	50	11.8	1404	61	11.7
BCRA	1407	48	11.7	1390	56	11.5
MaxRC-F	1394	49	11.6	1375	51	11.4
LIOA	1394	49	11.6	1375	52	11.4
ILIOA	1396	49	11.6	1374	51	11.4
Short lived connections						
Algorithm	$\lambda=2, T=20, \text{Length Run} = 50000$			$\lambda=3, T=20, \text{Length Run} = 50000$		
	Average number of Rejected LSPs	Standard deviation	% rejected LSPs	Average number of Rejected LSPs	Standard deviation	% rejected LSPs
MIRA	6200	176	12.4	12264	166	24.5
BU_MIRA	3849	129	7.6	10509	163	21.0
CSPF	3552	200	7.1	10330	156	20.6
BCRA	3617	151	7.2	10482	126	20.9
MaxRC-F	3556	144	7.1	10317	139	20.6
LIOA	3385	143	6.7	10262	167	20.5
ILIOA	3349	136	6.6	10196	168	20.3

5 Conclusions

The paper proposes ILIOA, which is an improved version of LIOA. The proposed algorithm have been compared with algorithms: MIRA, BU_MIRA, CSPF, MaxRC-MinF and BCRA. Algorithms Min Hop, Widest Shortest Path, which in the literature are often used only for comparative purposes, have been omitted.

The problem considered in this work is important not only in the MPLS network protocol, but also in other applications requiring dynamic allocation of network bandwidth, such as optical networks. Obtained results prove that in static conditions the number of rejected LSPs by ILIOA, LIOA and MaxRC-MinF is almost the same for networks of different sizes and different values of the load. However, in dynamic conditions of network performance ILIOA rejects the smallest number of requests, regardless of the load and size of the network.

The drawback of all considered algorithms is the treatment of weights of links independently. Therefore, further studies should lead to the introduction of the relationship between the weights of links in the network.

References

1. Bagula, A.B., Botha, M., Krzesiński, A.E.: Online Traffic Engineering: The Least Interference Optimization Algorithm. In: Proceeding of IEEE ICC 2004, pp. 1232–1236 (2004)
2. Guerin, R., Williams, D., Orda, A.: QoS Routing Mechanisms and OSPF Extensions. In: Proceedings of Globecom (1997)
3. Hodzic, H., Secerbegovic, E.: Online constraint -based routing as support for MPLS Traffic Engineering. In: 51st International Symposium ELMAR 2009, Zadra, Croatia (September 28-30, 2009)
4. Kodialam, M., Lakshman, T.V.: Minimum Interference Routing with Applications to MPLS Traffic Engineering. In: Proc. INFOCOM (2000)
5. Kotti, A., Hamza, R., Bouleimen, K.: Bandwidth Constrained Routing Algorithm for MPLS Traffic Engineering. In: Third International Conference on Networking and Services, ICNS (2007)
6. Krachodnok, P.: Constraint -Based Routing with Maximize Residual Bandwidth and Link Capacity-Minimize Total Flows Routing Algorithm for MPLS Networks. In: Fifth International Conference on Information, Communications and Signal Processing, pp. 1509–1514 (2005)
7. Olszewski, I.: Algorytmy wyboru ścieżek LSP w sieciach MPLS. *Kwartalnik Elektroniki i Telekomunikacji* 50(1), 7–23 (2004)
8. Zhu, M., Ye, W., Feng, S.: A new dynamic routing algorithm based on minimum interference in MPLS Networks. In: 4th International Conference on Wireless Communications, Networking and Mobile Computing, WICOM (2008)

Comparison of Modified Degree 6 Chordal Rings

Sławomir Bujnowski¹, Bożydar Dubalski¹, Antoni Zabłudowski¹,
Damian Ledźinski¹, Tomasz Marciniak¹, and Jens. M. Pedersen²

¹ University of Technology and Life Sciences,
Bydgoszcz, Poland

² Department of Electronic Systems,
Aalborg University, Aalborg, Denmark

Summary. In this paper we introduce a number of variants of modified degree 6 chordal rings, and we evaluate and compare their transmission properties in terms of average distance and diameter. We present theoretical models for calculating the distances, using optimal and ideal graphs, which are also shown to provide fairly good estimates. When comparing the distances, it turns out that the new suggestions for modification of the chordal rings results in lower distances, making them potentially interesting for use in communication networks. In the end of the paper we suggest directions for future research, in particular to investigate to what extent the topologies are suitable to implement.

1 Introduction

A fundamental task in designing telecommunications systems is choosing the interconnection network topology as it has a significant impact on efficiency, speed, and reliability of the entire system [1]. Moreover, the topology is a fundamental part of the network and linked to the physical implementation which is difficult and expensive to change once the network is established: It is therefore important to carefully choose the best possible topology in terms of both performance and cost. Nowadays, analysis of regular network structures is one of the most important issues in telecommunications and computer science. Using such topologies makes it possible to reduce the complexity of routing and restoration schemes, making network planning and management easier. It also leads to networks properties which are easier to predict, also in case of network failures. In this paper we expand previous work on chordal rings and modified chordal rings in order to introduce a number of modifications to chordal rings of degree 6, and we show that their performance in terms of average distance and diameter is better than the original chordal rings, and also better than previously defined modified chordal rings. The

paper is organized as follows. Section 2 gives the background definitions, and the well know chordal rings and modified chordal rings of degree 6 are presented along with our two proposed modifications. In Section 3 we introduce the optimal and ideal graphs, and show how these can be used to estimate the distance parameters in the proposed graphs. In Section 4 we compare the different topologies, and validate that the theoretical estimates of distances are quite precise. Section 5 contains conclusions and directions for further research.

2 Background

The networks can be model by symmetric digraphs, i.e., a directed graph G with vertex set $V(G)$ and edge set $E(G)$, such that, if $[vi, vj]$ is in $E(G)$, then $[vj, vi]$ is also in $E(G)$ [2]. Due to the symmetry, we will in this paper simply model the networks by undirected graphs. This also matches the fact that most optical networks are constructed by two rings (one in each direction) running in the same trunk. Among the analyzed topologies that would be used in designing the distributed structures, the authors have chosen rings as they are very simple and extensible.

In [3, 4, 5] the authors analysed the transmission properties of third, fourth and sixth degree standard chordal rings. In the [6, 7, 8] modifications of double rings structures N2R and third and four degree chordal rings were proposed. In all these cases the graphs possess two different chord lengths, and the obtained results of the calculation of two basic parameters of the graphs - diameter and average distance - are much shorter than in the standard graphs. Similar results have been obtained for sixth degree chordal rings [9], but the analysis presented only concerned a subset of these graphs. For this reason the authors decided to inspect sixth degree chordal rings with different chord lengths and to compare the obtained results with parameters of other ring structures of the same degree.

Definition 5 *A chordal ring is a ring with additional arcs called chords. The chordal ring is defined by the pair (p, Q) , where p denotes the number of nodes appearing in the ring and Q denotes the set of chord lengths $Q \subseteq \{2, \dots, \lfloor p/2 \rfloor\}$. Each chord of length $q_i \in Q$ connects every two nodes of the ring that are at distance q_i . The chordal ring is denoted as $CHRd(V)(p; 1, q_2, \dots, q_j)$, $q_2 < \dots, q_j$. In general, the nodal degree of chordal rings $d(V) = 2j$, unless there is a chord of length $p/2$, in this case p should be even and nodal degree equal to $2i - 1$ [10].*

An example of such a graph is shown in Fig. 11

The ideas and results of modified degree 6 chordal rings were presented previously [9, 10].

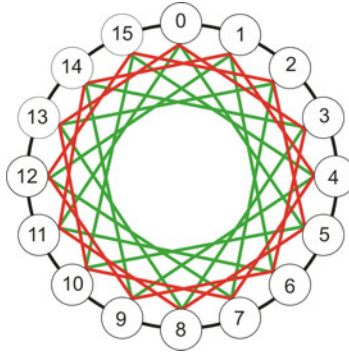


Fig. 1. Chordal ring of 6th nodal degree CHR6(16; 3,5)

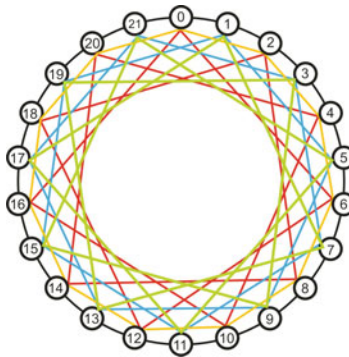


Fig. 2. Modified chordal ring CHR6m(22; 1,2,4,6)

Definition 6 *The modified degree six chordal ring called $CHRM6$ is an undirected circulant graph. It is denoted by $CHRM6(p; 1, q_1, q_2, q_3)$ where p is the positive and even number of node, $q_1, q_2,$ and q_3 are chords even length ($q_1 < q_2 < q_3$). $CHRM6$ consists of one ring connecting p nodes. Each even node i_{2k} and odd node and i_{2k+1} ($0 \leq k < p/2$) is additionally connected to four nodes: i_{2k} is connected to $i_{2k-h_1(mod p)}, i_{2k+h_1(mod p)}, i_{2k-h_3(mod p)}, i_{2k+h_3(mod p)}$, while i_{2k+1} is connected $i_{2k+1-h_2(mod p)}, i_{2k+1+h_2(mod p)}, i_{2k+1-h_3(mod p)}, i_{2k+1+h_3(mod p)}$. The values of p, q_1, q_2, q_3 must fulfill condition $GCD(p, q_1, q_2, q_3) = 2$ [9].*

The obtained results show that this type of structure has much better parameters (diameter and average path length) than original (standard) chordal ring sixth nodal degree, which inspired the authors to examine two other modified topologies.

3 Other Modified Topologies of Chordal Rings 6th Nodal Dergree

The modified chordal ring suggested in [9] is a special case of the chordal ring, which we generalize in definition 3.

Definition 7 *The modified degree six chordal ring called CHRma6 is an undirected circulant graph. It is denoted by CHRma6(p, q₁, q₂, q₃, q₄) where p is the positive and even number of nodes; q₁, q₂, q₃ and q₄ are chords of even length. Each even node i_{2k} is connected to six other nodes: i_{2k-1}, i_{2k+1}, i_{2k-h1(mod p)}, i_{2k+h1(mod p)}, i_{2k-h2(mod p)}, i_{2k+h2(mod p)}, while odd node i_{2k+1} is connected to i_{2k}, i_{2k+2}, i_{2k+1-h3(mod p)}, i_{2k+1+h3(mod p)}, i_{2k+1-h4(mod p)}, i_{2k+1+h4(mod p)} (0 ≤ k < p/2). The values of p, q₁, q₂, q₃, q₄ must fulfill condition GCD(p, q₁, q₂, q₃) = 2.*

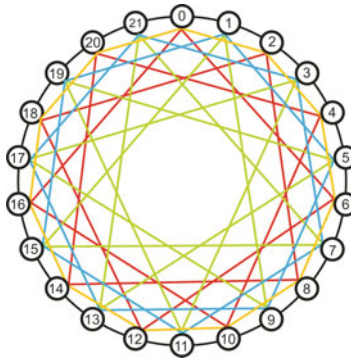


Fig. 3. Modified chordal ring CHR6ma(22; 2,4,6,8)

In order to perform the analysis, two special types of chordal rings will be defined. The first one is called the **ideal chordal ring (graph)** and the second one - the **optimal chordal ring**. Their parameters will be used as the reference for evaluation of the parameters of examined structures.

Definition 8 *An ideal chordal ring is a regular graph with the total number of nodes p_i expressed by the formula:*

$$p_i = 1 + \sum_{d=1}^{d(G)-1} |p_d| + |p_{d(G)}| \tag{1}$$

where p_d denotes the number of nodes belonging to the d – th layer (the layer means the subset of nodes that can be reached from any source node using d edges), and p_{d(G)} the number of nodes appearing in the last layer. In the ideal rings, for every d < d(G), the number of nodes p_d reaches maximal value,

for each $j \neq k < d(G)$ $p_j \cap p_k = \emptyset$ and all the remaining nodes appeared in layer $p_{d(G)}$. If the subset $p_{d(G)}$ reaches the maximal possible value, then such a ring is called the optimal chordal ring.

In order to begin the analysis of this proposed topology optimal graphs were determined. In Table 1 the maximal numbers of nodes which are equally distant from an arbitrary source node in the (virtual) modified optimal graphs are shown.

Table 1. Maximal number of nodes in the layers

d	1	2	3	4	5	6	7	8
p_d	6	26	98	298	738	1562	2946	5098

d denotes the distance, and p_d the number of nodes appearing in $d - th$ layer.

If $d > 2$ the power of these sets is described by formula:

$$p_d = \frac{4}{3}d^4 - \frac{28}{3}d^2 + 32d - 22 \tag{2}$$

On the basis of the formula given above the total number of nodes p_o forming the optimal graph with diameter $d(G)$ can be calculated ($d(G) > 1$):

$$p_o = \frac{4}{15}d(G)^5 + \frac{2}{3}d(G)^4 - \frac{8}{3}d(G)^3 + \frac{34}{3}d(G)^2 - \frac{38}{5}d(G) + 5 \tag{3}$$

The total numbers of nodes forming optimal graphs versus diameter is shown in Table 2.

Table 2. Total numbers of nodes forming optimal graphs versus diameter

$d(G)$	1	2	3	4	5	6	7	8
p_o	7	33	131	429	1167	2729	5675	10773

The average path length in the optimal graphs can be calculated using formula:

$$d_{avo} = \frac{\sum_{d=1}^{d(G)} dp_d}{p_o - 1} = \frac{5}{6} \frac{2d(G)^6 + 6d(G)^5 - 16d(G)^4 + 54d(G)^3 + 23d(G)^2 - 47d(G) + 36}{2d(G)^5 + 5d(G)^4 - 20d(G)^3 + 85d(G)^2 - 57d(G) + 30} \tag{4}$$

If $d(G) \gg 1$:

$$d_{avo} \cong \frac{5}{6} (d(G) + 0, 5) \tag{5}$$

Table 3. Examples of ideal graphs

p	$d(G)$	d_{av}	q_{1e}	q_{1o}	q_{2o}	q_{2o}
26	2	1,760	4	6	8	12
38	3	1,973	4	16	8	12
44	3	2,116	6	10	14	18
46	3	2,156	4	16	8	12
50	3	2,225	4	14	8	12
52	3	2,255	6	14	10	18
54	3	2,283	4	14	8	20
56	3	2,309	6	10	18	26
58	3	2,333	4	14	8	18
62	3	2,377	4	14	8	28
64	3	2,397	6	22	14	30

which allows us to estimate the expected value of this parameter without complicated calculations.

The optimal graphs are virtual - they don't exist in reality - but the ideal rings can be found. Examples of these graphs are shown in Table 3.

In Fig. 4 the diameter and average distances of the best modified chordal rings as a function of the numbers of nodes are shown.

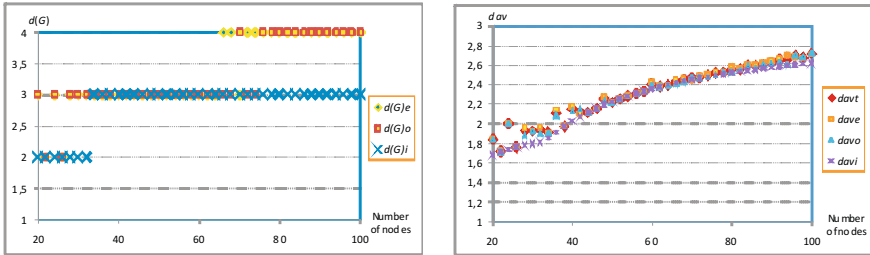


Fig. 4. Charts of diameter and average path length in function of node number. Index e means even number of node, o - odd number, i - ideal, t - average value

Another possible chordal ring of degree six is described as follows.

Definition 9 The modified degree six chordal ring called $CHRmh6$ is denoted by $CHRmh6(p; q_1, q_2, q_3)$ where p is the positive and even number of nodes; q_1, q_2 and q_3 are chords, where chord q_1 generates a Hamiltonian cycle, and q_2 and q_3 possess even lengths. Each even node i_{2k} is connected to six other nodes: $i_{2k-1}, i_{2k+1}, i_{2k-h1(mod p)}, i_{2k+h1(mod p)}, i_{2k-h2(mod p)}, i_{2k+h2(mod p)}$, while an odd node i_{2k+1} is connected to $i_{2k}, i_{2k+2}, i_{2k+1-h3(mod p)}, i_{2k+1+h3(mod p)}, i_{2k+1-h3(mod p)}, i_{2k+1+h3(mod p)}$ ($0 \leq k < p/2$). The values of p and q_1 must be prime to each other, whereas q_2 and q_3 must fulfill that $GCD(p, q_2, q_3) = 2$.

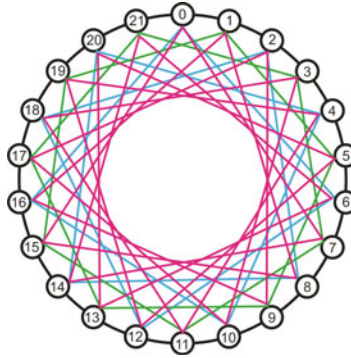


Fig. 5. Modified chordal ring CHR6mh(22; 7,4,6)

In this case optimal the maximal numbers of nodes in the consecutive layers is shown in Table 4

Table 4. Maximal number of nodes in the layers

d	1	2	3	4	5	6	7	8
p_d	6	26	80	184	352	600	944	1400

d denotes the distance, and p_d the number of nodes appearing in d -th layer. If $d > 2$ the power of these sets is described by formula:

$$p_d = \frac{7}{3}d^3 + \frac{16}{3}d - 8 \tag{6}$$

Using the formula given above the total number of nodes p_o in the optimal graph with diameter $d(G)$ was calculated ($d(G) > 1$):

$$p_o = \frac{2}{3}d(G)^4 + \frac{4}{3}d(G)^3 + \frac{10}{3}d(G)^2 + \frac{16}{3}d(G) + 9 \tag{7}$$

Table 5. Total numbers of nodes forming optimal graphs versus diameter

$d(G)$	1	2	3	4	5	6	7	8
p_o	7	33	113	297	649	1249	2193	3593

The average distance in optimal graphs is describing by formula:

$$\begin{aligned} d_{avo} &= \frac{\sum_{d=1}^{d(G)} dp_d}{p_o - 1} = \frac{6 + \frac{8}{15}d(G)^5 + \frac{4}{3}d(G)^4 + \frac{8}{3}d(G)^3 - \frac{4}{3}d(G)^2 - \frac{16}{3}d(G) + 10}{\frac{2}{3}d(G)^4 + \frac{4}{3}d(G)^3 + \frac{10}{3}d(G)^2 + \frac{16}{3}d(G) + 8} = \\ &= \frac{\frac{4}{15}(2d(G)^5 + 5d(G)^4 + 10d(G)^3 - 20d(G)^2 - 12d(G) + 60)}{\frac{2}{3}(d(G)^4 + 2d(G)^3 + 5d(G)^2 + 8d(G) + 12)} = \\ &= \frac{2}{5} \frac{2d(G)^5 + 5d(G)^4 + 10d(G)^3 - 20d(G)^2 - 12d(G) + 60}{d(G)^4 + 2d(G)^3 + 5d(G)^2 + 8d(G) + 12} \end{aligned} \tag{8}$$

When $d(G) \gg 1$:

$$d_{avo} \cong \frac{2}{5} (2d(G) + 1) \tag{9}$$

In Fig. 6 the diagrams of real and ideal chordal rings parameters are shown. From these diagrams it follows that there exist many ideal graphs (see Table 6).

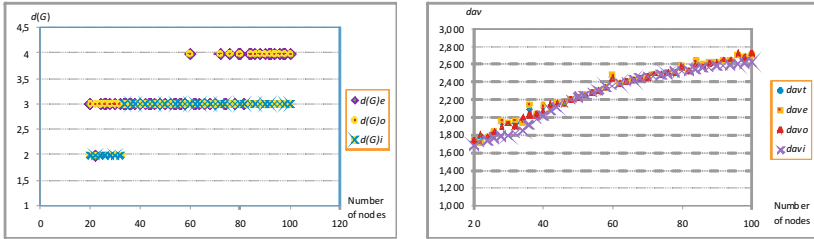


Fig. 6. Charts of diameter and average path length in function of node number. Index e means even number of node, o - odd number, i - ideal, t - average value.

Table 6. Examples of ideal graphs

p	$d(G)$	dav	q_1	q_{2e}	q_{2o}
46	3	2,156	5	8	16
48	3	2,191	7	10	22
50	3	2,224	7	12	16
52	3	2,255	5	14	22
54	3	2,283	5	8	20
56	3	2,309	9	6	26
58	3	2,333	5	8	20
64	3	2,397	5	14	26
66	3	2,415	5	8	26
68	3	2,433	5	18	30
70	3	2,449	17	4	24
74	3	2,479	5	18	30
78	3	2,506	19	4	28
82	3	2,531	21	8	24

4 Comparison of Sixth Nodal Degree Chordal Rings

To sum up, Fig. 7 shows the comparison of theoretical diameter and average path length appearing in ideal and optimal degree 6 chordal rings as a function of the number of nodes. It can be seen that the newly proposed modified chordal rings have shorter distances than the original chordal rings, and for large graphs also shorter distances than the previously suggested modified chordal rings.

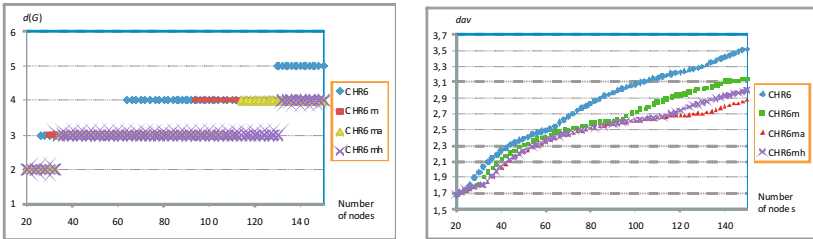


Fig. 7. Comparison of the theoretical diameter and average path length appearing in different types of chordal rings CHR6 versus number of nodes

In Fig. 8 the comparison of diameter and average distances appearing in real, best chordal rings of each type as a function of the number of nodes are shown. The tendencies are similar to theoretical results in Fig. 7, and also confirm that the theoretical estimates are fairly precise.

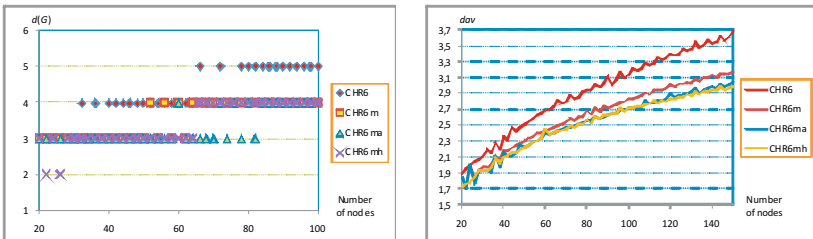


Fig. 8. Comparison of the diameter and average path length appearing in different types of real chordal rings CHR6 versus number of nodes

In Fig. 9 the theoretical and real difference of average distances between analyzed graphs in relation to CHR6ma, expressed in percents, is shown.

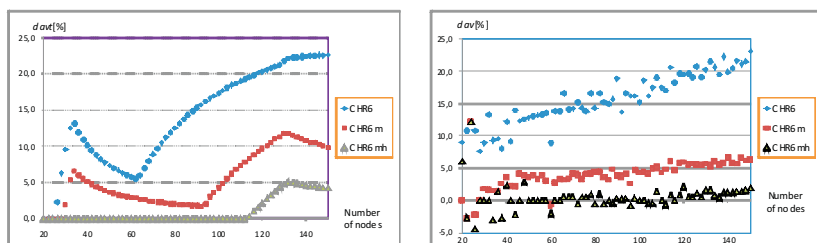


Fig. 9. Comparison of average distance in CHRm and CHR6th versus number of nodes. The left diagram presents the theoretical values, and the right diagram the real obtained values.

5 Conclusions

This paper has introduced two new kinds of modified chordal rings of degree 6, and compared the average distance and diameter of these with standard chordal rings as well as with previously introduced modified chordal rings. Both average distance and diameter turns out to be lower in the new modified chordal rings, which indicates that they have superior transmission properties. The results are obtained using both practical evaluations on a large number of graphs, and by developing a theoretical model based on optimal and ideal graphs. It is shown that the theoretical model produces a fairly good estimation of the real values. In relative numbers, the average distance of the best modified chordal rings is around 5% shorter than the previously introduced chordal rings, and 20-25% shorter than the standard chordal rings.

Further research is needed in order to verify the applicability of these graphs in real networks, e.g. on the optical layer in Fiber networks. In practice, it is difficult to generally state that one topology is superior to another, since it depends on concrete requirements that differ from case to case. A starting point for further research could be to see how comparable topologies perform in a number of case studies. Another direction for further research could be to investigate the reliability (for example, the ability to offer short, disjoint paths) of the different topologies.

References

1. Bhuyan, L.N.: Interconnection Networks for Parallel and Distributed Processing. *IEEE Computer* 20(6), 9–12 (1987)
2. Narayanan, L., Opatrny, J., Sotteau, D.: All-To-All Optical Routing in Chordal Rings of Degree 4. *Algorithmica* 31, 155–178 (2001)
3. Bujnowski, S., Dubalski, B., Zabłudowski, A.: Analysis of Chordal Rings, Mathematical Techniques and Problems in Telecommunications. In: *Centro International de Matematica, Tomar 2003*, pp. 257–279 (2003)

4. Bujnowski, S., Dubalski, B., Zabłudowski, A.: The Evaluation of Transmission Ability of 3rd Degree Chordal Rings with The Use of Adjacent Matrix. In: Bujnowski, S., Dubalski, B., Zabłudowski, A. (eds.) The Seventh INFORMS Telecommunications Conference, Boca Raton, pp. 219–221 (2004)
5. Bujnowski, S., Dubalski, B., Zabłudowski, A.: Analysis of 4th Degree Chordal Rings. In: Proceedings of International Conference on the Communications in Computing, Las Vegas, pp. 318–324 (2004)
6. Dubalski, B., Bujnowski, S., Zabłudowski, A., Pedersen, J.: Introducing Modified Degree 4 Chordal Rings with Two Chord Lengths. In: Proceedings of the Fourth IASTED Asian Conference, Communication Systems and Networks, Phuket (2007) ISBN CD: 978-0-88986-658-4, 561–174–2
7. Bujnowski, S., Dubalski, B., Pedersen, J.M., Zabłudowski, A.: Struktury topologiczne CR3m oraz NdRm. Przegląd Telekomunikacyjny LXXXI(8/9), 1133–1141 (2008)
8. Dubalski, B., Zabłudowski, A., Bujnowski, S., Pedersen, J.M.: Comparison of Modified Chordal Rings Fourth Degree to Chordal Rings Sixth Degree. In: Proceedings of Electronics in Marine, ELMAR 2008, Zadar, Croatia, vol. 2, pp. 597–600 (2008)
9. Azura, R., Othman, M., Selamat, H., Hock, P.: Modified Degree Six Chordal Rings Network Topology, Prosiding Simposium Kebangsaan Sains Matematik Ke-16, 3–5 (2008)
10. Gavoille, C.: A Survey on Internal Routing, <http://deptinfo.labri.ubordeaux.fr/gavoille/article/survey/node28.html>
11. Kotsis, G.: Interconnection Topologies and Routing for Parallel Processing Systems. ACPC, Technical Report Series, ACPC/TR92-19 (1992)

Evaluation of Measurement Based Admission Control Algorithms for IEEE 802.16 Networks in Simulations with L2S Physical Layer Abstraction and nbLDPC Codes

Adam Flizikowski¹, Marcin Przybyszewski², and Witold Hołubowicz³

¹ University of Technology and Life Sciences,
Bydgoszcz, Poland

e-mail: adamfli@utp.edu.pl

² ITTI Ltd, Poznan, Poland

marcin.przybyszewski@itti.com.pl

³ Adam Mickiewicz University, Poznań, Poland

e-mail: holubowicz@amu.edu.pl

Summary. In this paper we present results of evaluating and validating the three novel measurement based admission control algorithms for WiMAX dealing with connection requests that arrive in batches (ARAC, nscARAC, EMAC). We focus on the capacity of a future 4G system using non binary LDPC codes. Simulations rely solely on the use of VoIP traffic flows (with and without voice detection). In order to avoid actual FEC decoder implementation we use L2S interface described in IEEE 802.16 EMD. We use performance metrics characteristic to admission control to assess the system performance. Simulated nodes move according to the Leavy Walk distribution. All algorithms presented in the paper have been simulated using the ns2 platform with ViMACCS extension developed by authors as part of previous work.

1 Introduction

Admission Control in WIMAX has been a hot research topic since the introduction of IEEE802.16-2005. The new standard had indicated the need to implement admission control, but did not define any specific algorithm, leaving this task open to vendors. Number of algorithms have been proposed and evaluated through analytical models. In this paper we intent to assess performance of various measurement based admission control algorithms using our WiMAX patch for ns2, namely ViMACCS [2, 3, 4, 5]. The ultimate goal behind the research carried by authors is to assess the QoE of VoIP especially in system under tests implementing the nb-LDPC codes developed by DaVinci project [6]. Thus we are interested in VoIP performance

evaluation under various system conditions (e.g. mobility in different geographical areas). Since ViMACCS is based on NIST WiMAX module extended by A. Belight [7] we treat VoIP with silence suppression as rtPS, since currently there is no support for ertPS service flow class.

This paper is organized as follows: in Section 2 a brief description of our previous work is given. Section 3 describes two admission control (MBAC) algorithms - Arrival Rate - based Admission Control (ARAC) and Arrival Rate - based Admission Control with no connection state control (nscARAC). Also in this section we present results obtained for both algorithms. Section 4 presents results obtained for the ARAC and simple algorithm utilizing exponentially moving average (EMA), which has been originally presented in [8] (in this article referred to as EMAC). We conclude with Section 5, where we present plans for our future work.

2 Previous Work

In order to assess performance improvements introduced by 4G-candidate systems, a comprehensive link level, cell level and eventually system level simulations are required. The need to evaluate overall system performance that has been enabled by the cooperation of multiple performance enhancing techniques like MIMO, novel channel coding schemes like non-binary LDPC codes together with standards like IEEE 802.21, which aim at enabling handover and interoperability between heterogeneous network technologies, make system level simulations even more important. Therefore the simulation environment providing both MAC and physical layer realizations at the satisfactory level of fidelity and standard compliance is needed. Our previous work had concentrated on selecting, developing and improving WiMAX module for ns2 simulator [2, 3, 4, 5]. Authors have successfully tested different approaches to simulation environment with the core idea of integrating the ns2 with variety of physical layer abstraction implementations. Resource management research is part of our previous work on the admission control strategies that were successfully adopted, adjusted and evaluated to show system performance under various load and (guard-channel based) admission control algorithms configurations [2, 3, 4, 5]. The admission control performance was assessed assuming different FEC schemes introduced in a physical layer (e.g. CTC, BTC, LDPC, nbLDPC). Also in [5] we have tested different approaches to parameter based admission control, congestion control and symbol reservation schemes in case of quasi - mobility [5]. In [5] we introduced an approach utilizing maps of geographical area with SNR distribution calculated. In the following paper a Link to System (L2S) interface is adopted in order to obtain more accurate results (as compared to previously used Gaussian noise channel).

3 ARAC with MCS and Connection State Control (ARAC) and ARAC Without MCS and Connection State Control (nscARAC)

The purpose of this section is to compare performance of the two novel measurement based (MBAC) algorithms, namely:

- ARAC with MCS and connection state control (referred to simply as "ARAC")
- ARAC without MCS and connection state control (nscARAC - no state control ARAC).

In related simulations scenarios we compare the two algorithms for different configuration of arrival rate, average connection holding time and resource utilization averaging interval (in frames).

3.1 Introduction

One of the compared algorithms nscARAC takes an advantage of the fact that Base Station (BS) has the ability to monitor information about current arrival rate. Based on this value BS estimates, using moving average statistics of resource usage, takes into consideration recently accepted connections. If connection requests start arriving in large batches, than in order to estimate value of available (OFDM) symbols, nscARAC takes into consideration QoS parameters (e.g. MSTR) of connections that have already been accepted but do not exist long enough to heavily influence the calculation of an average symbol utilization (Fig. 1). As a part of algorithm we also introduce notion of guard channel (GC) threshold for VBR traffic.

The extended version of nscARAC was introduced with controlling resource usage due to MCS (modulation and coding scheme) and connection state control - ARAC. ARAC not only measures arrival rate, but (in the

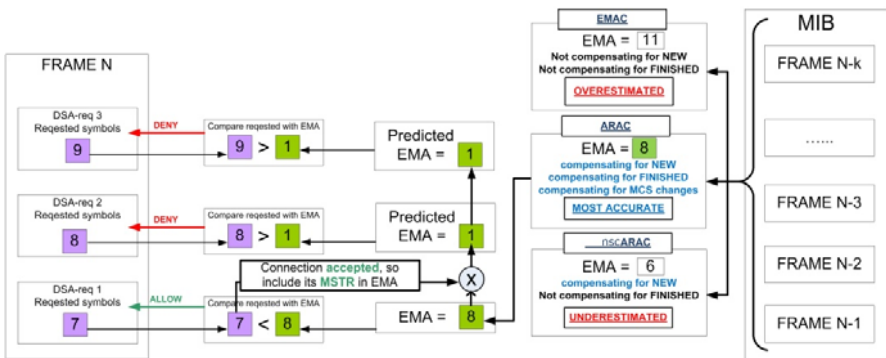


Fig. 1. Comparison of ARAC, nscARAC and EMAC algorithms

process of calculating EMA) it also improves the EMA calculation by excluding connections which do not exist anymore although their past statistics still influence EMA. Moreover ARAC compensates for connections that have changed their MCS during connection, therefore their current bandwidth requirements have changed. ARAC algorithm has been presented in Fig. 4 and Fig. 6. The number of frames/time required to compute EMA is referred to as "averaging interval".

3.2 Simulation Parameters and Results

There are two scenarios - scenario S1 and S2. The value of arrival rate for both scenarios is changing in subsequent simulations and is in range from 25 to 250 connections per minute per Service Flow class (e.g. 250 per minute for UGS and 250 per minute for rtPS). All users are moving along geographical area represented by maps described by SNR distribution (Map 1). In order to be able to measure performance of the tested MBAC algorithms for different traffic types, the CBR VoIP and VBR VoIP traffic with silence suppression were introduced. VBR connections are scheduled as rtPS traffic. For VBR rtPS VoIP two codecs were used namely the G.711 and G.729 with "one-to-one" voice detection model. The amount of simulated nodes complying with the two types of VoIP traffic is exactly the same (e.g. 10 users with G.711, 10 with G.729 and 10 with CBR). In order to generate real-life VoIP traces we have integrated the extended ns2 VoIP traffic generators developed as part of the FP6 EuQoS european project [10] into our simulator (ViMACCS) [2, 3, 4, 5]. At the physical layer the nb-LDPC (DaVinci) FEC scheme is used for all simulations. Admission Control algorithm is also used as a congestion control algorithm. Guard channel for VBR traffic is set to 10% for all simulations. The core network delay is fixed with value of 50ms. We assume that QoS of particular connection has not been met once its end-to-end delay is greater than 150ms (a maximum of 100ms in the access plane). Simulation parameters for scenario S1 and S2 have been presented in Table 1 below.

Scenario S1 results. In this scenario we test AC performance for short averaging intervals and medium connection holding times. Therefore in scenario S1 the averaging interval is set to value of 0.4 seconds (20 frames) and average connection holding time is set to 20 seconds. Results obtained for this scenario are presented below. All the figures show the average values together with 95% confidence interval (outliers in the charts).

Scenario S2 results. In this scenario we test AC performance for long averaging intervals and short connection holding times. Therefore in scenario S2 averaging interval is set to 4.0 seconds (200 frames) and average connection holding time is set to only 2 seconds.

Table 1. Simulation parameters for scenarios S1 and S2

Network configuration parameters	Value
SF class	UGS rtPS
CBR VoIP (UGS)	64 kbps
VBR VoIP codecs (rtPS)	G.711 G.729
VBR VoIP Voice Detection Model	One-to-one
Link2System (L2S) abstraction	RBIR
Simulation time	200s
Congestion Control	Enabled
Scenario Repetitions	6
Coding scheme	nb-LDPC (DaVinci)
Target CWER	0.01
Carrier frequency	3.5 GHz
Bandwidth	3.5 MHz
Number of sub-carriers	256
Number of data sub-carriers	192
Cyclic prefix	1/8
Modulation	QPSK, 16-QAM, 64-QAM
Codeword length	48, 96, 144, 288
Rates	1, 2/3, 3, 5/6
Mobile node average speed	0.83 m/s
Transmission direction	Uplink

Table 2. Traffic characteristics

Network configuration parameters	Value
Arrival rate (per Service Flow)	25 to 250(conn/min)
Average Connection Time	20s (S1) 2s (S2) - exponential
CAC	nscARAC ARAC
Averaging Interval	0.4s (S1) 4.0s (S2)

3.3 Results Discussion and Conclusions

For scenario S1 (Fig. 2, Fig. 3) it can be observed that there is no difference in performance between ARAC and nscARAC. That is because for short averaging intervals the small arrival rates and medium average connection times difference in EMA estimation between ARAC and nscARAC is not high enough to make a difference in blocking or dropping probability. This difference can clearly be observed for scenario S2. For high averaging interval (200 frames), low average connection holding times (2s) and high arrival rates ARAC achieves blocking probabilities lower by almost 10% (Fig. 4) at the same time keeping dropping probabilities at the same level (Fig. 5). It is worth noticing that both algorithms independent on the scenario succeed with provisioning the QoS for all admitted connections (see Table 3).

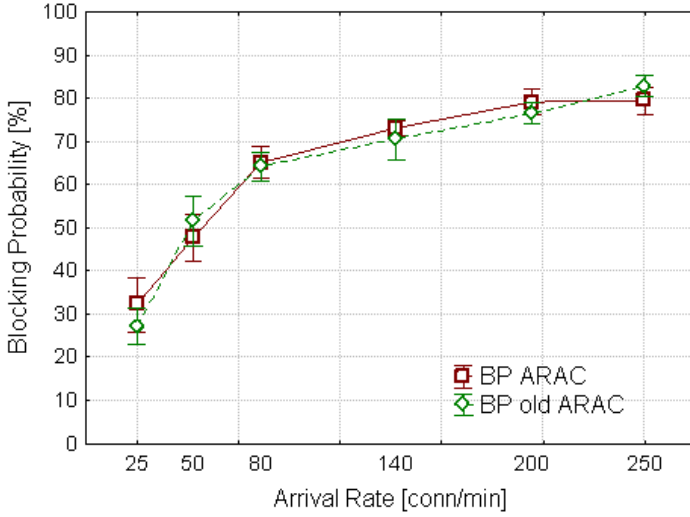


Fig. 2. Blocking probabilities for ARAC and nscARAC - scenario S1

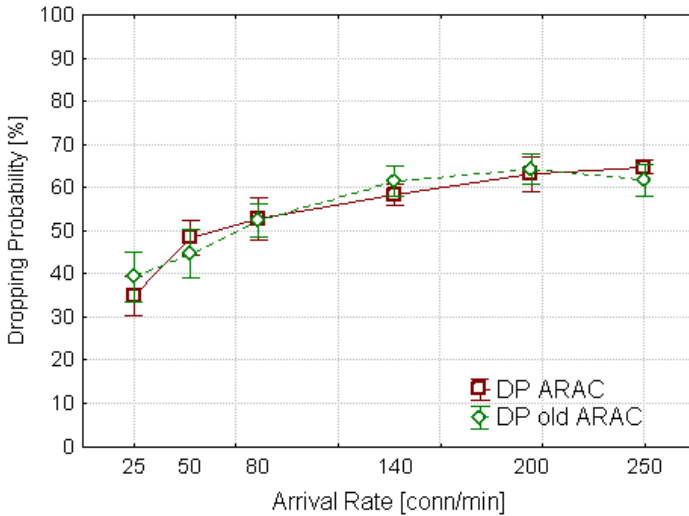


Fig. 3. Dropping probabilities for ARAC and nscARAC - scenario S1

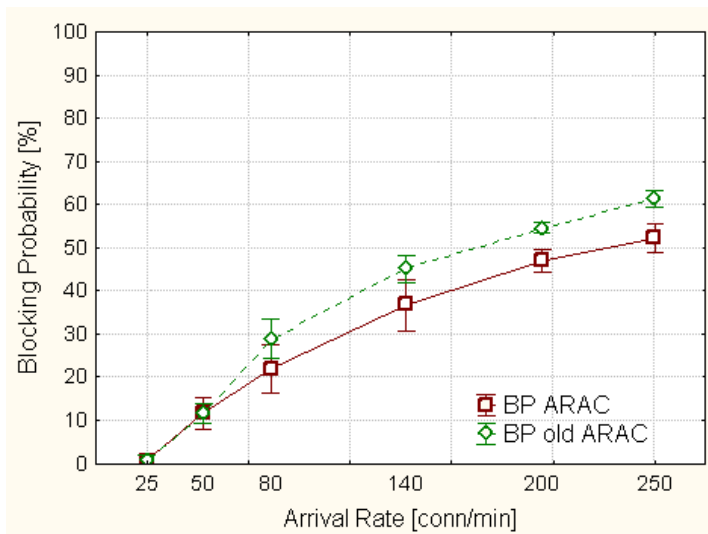


Fig. 4. Blocking probabilities for ARAC and nscARAC S2

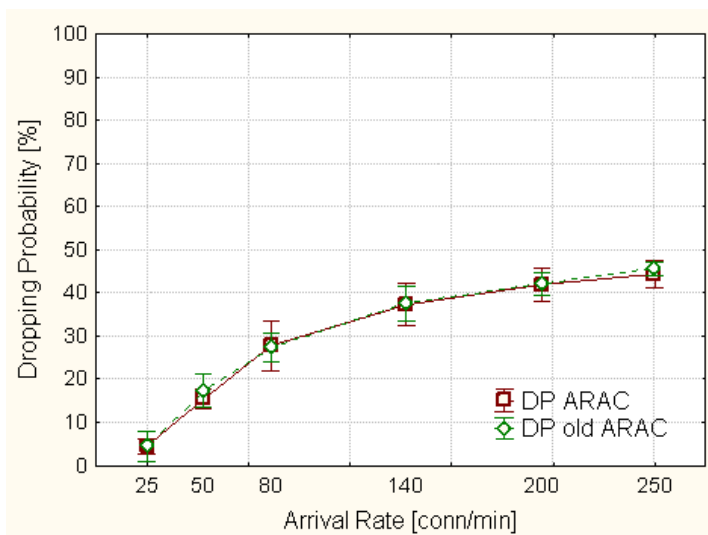


Fig. 5. Dropping probabilities for ARAC and nscARAC S2

Table 3. Simulations results for arrival rate $\lambda = 280$ [conn/min]

Scenario	S1		S2		S3 ¹	
AC algorithm	ARAC	nscARAC	ARAC	nscARAC	ARAC	EMAC
Percent of						
connections with	100	100	100	100	100 ¹	52 ¹
satisfied QoS [%]						
Blocking Probabil. [%]	72	70	38	47	65 ¹	65 ¹
Dropping Probabil. [%]	59	61	38	38	62 ¹	38 ¹

¹for averaging interval 500 frames

Conclusions. For long averaging intervals and short connection holding times the ARAC offers lower blocking probabilities than nscARAC (Fig. 4). Nevertheless ARAC is a more complex solution as it has to monitor also MCS changes and take into consideration already finished connections. Therefore the longer averaging interval, the more computationally complex ARAC algorithm becomes. This, together with the fact that both ARAC and nscARAC succeed to provide an adequate QoS level for all ongoing connections, makes ARAC a better choice only for systems with long averaging interval and where ARAC complexity will not be an issue (but this issue could be solved by implementing a robust MiB).

4 Performance Comparison of ARAC and EMAC

The purpose of this section is to compare performance of two measurement based admission control algorithms (MBCAC or MBAC) - namely ARAC and EMAC. In simulations for this scenario we change the number of frames used to calculate EMA (Exponential Moving Average) for a constant (high) arrival rate.

4.1 Introduction

One of the compared MBAC algorithms is a simple to implement algorithm utilizing exponentially moving average (EMA) which has been presented in [8] (in the this article referred to as EMAC). Since EMAC does not provide protection against the problem of appropriate estimation of resources when connection requests start arriving in large batches (the EMAC underestimates number of used symbols - Fig. 6). The other algorithm is the ARAC introduced in previous section. By taking into consideration parameters of the connections that do not exist long enough to influence EMA in the admission process and compensating for connections that have changed their MCS, and compensating for connections that do not exist anymore, but still influence EMA, ARAC should provide means to cope with the both the problem of over- and underestimation of free resources.

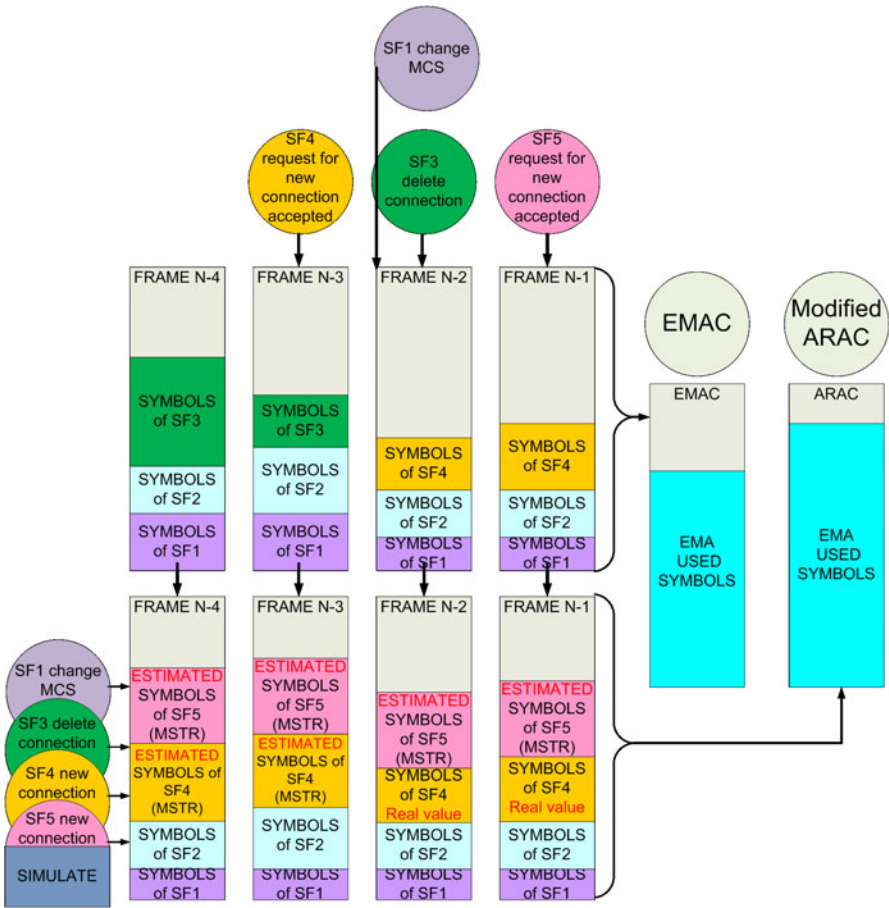


Fig. 6. EMAC and "old" ARAC \tilde{U} example of estimating resources for $k = 4$ frames

4.2 Simulation Parameters and Results

Within the set of simulated mobile nodes there are 60% of users that utilize codecs with silence suppression. The level of arrival rate is set to 140 connections per minute per Service Flow class (140 per minute for UGS and 140 per minute for rtPS). We assume the core network delay to be at a constant level of 50ms. We assume, that QoS of particular connection has not been met if its end-to-end delay is greater than 150ms (100ms in the access plane). The rest of the parameters have been set according to the values given in the previous section.

4.3 Results Discussion and Conclusions

Results discussion. As it can be seen for short and medium averaging intervals the EMAC algorithm offers smaller blocking probabilities (Fig. 7) and similar dropping probabilities (Fig. 8) than ARAC. This is because by

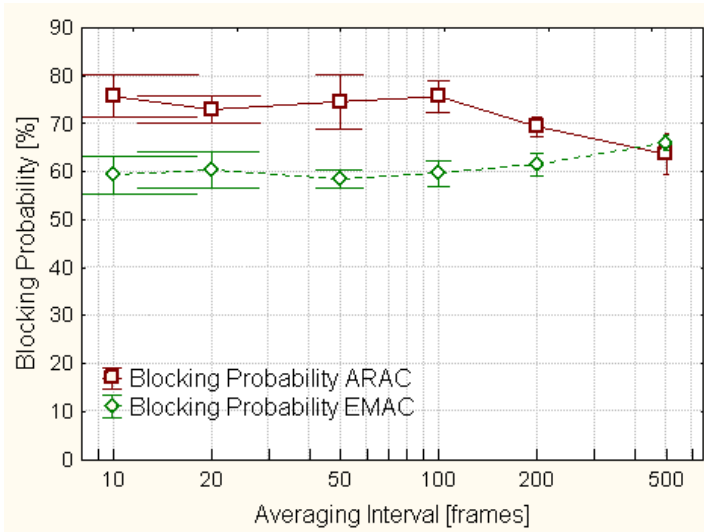


Fig. 7. Blocking probabilities for ARAC and EMAC

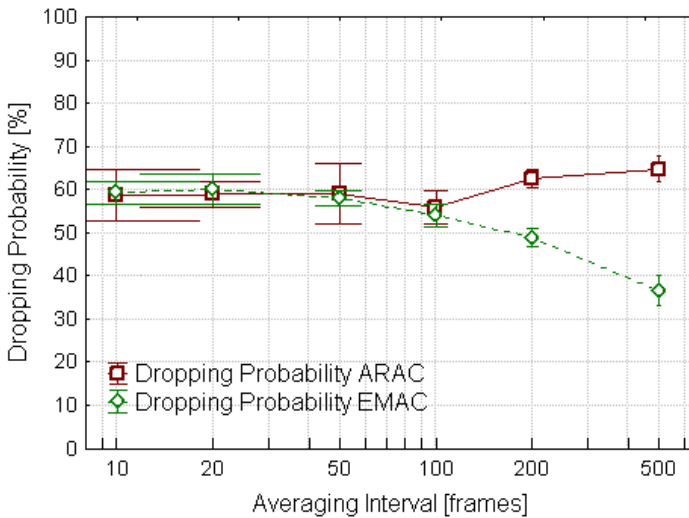


Fig. 8. Dropping probabilities for ARAC and EMAC

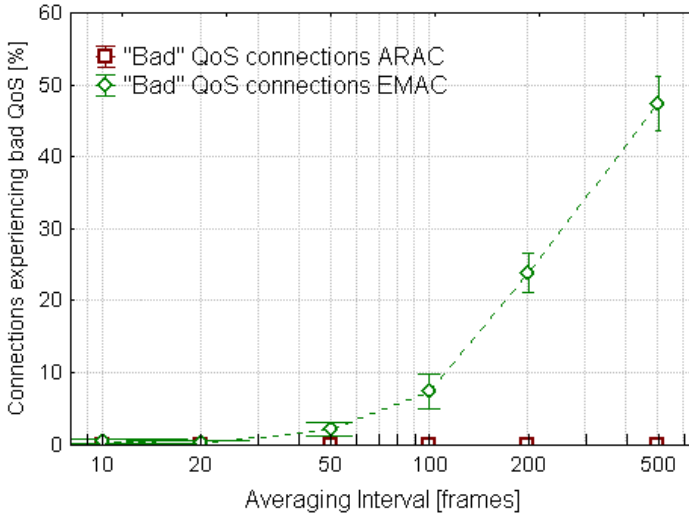


Fig. 9. Number of connections with "bad" QoS for ARAC and EMAC

considering the average utilization of resource consumed by newly accepted connections, for the same system conditions, ARAC's moving average is lower than the same statistics calculated for EMAC (assuming that EMA statistics is a measure of free resources). Nevertheless, for high averaging intervals we can observe ARAC begins to outperform EMAC. This is due to EMAC errors in estimations of EMA. EMAC's average for long averaging intervals does not reflect actual resource's current state at BS. Figure Fig. 9 presents percentage of connections for which QoS requirements have not been met. For ARAC the QoS requirements for all admitted connections have been met for all simulated cases. On the other hand EMAC fails to provide QoS for all connections. The higher the number of frames is used to calculate EMA (or in other words - for higher averaging intervals) the higher the percentage of connections with "bad" QoS. For long averaging intervals (e.g. 500 frames) almost half of the connections do not meet their QoS requirements. By choosing shorter averaging intervals number of "bad" QoS connections can be minimized. But even for very short simulation times (e.g. 10 subsequent OFDM frames which is equal to 200 milliseconds in time domain) EMAC fails to provision QoS to all connections independently of the averaging period - about 0.9% of connections experience bad QoS.

Conclusions. Although EMAC generally offers lower blocking probabilities than ARAC (which can be considered one of major performance factors for CAC algorithm) it cannot be considered a more robust algorithm. The fact that EMAC is not able to provide an appropriate QoS level for connections that have been admitted, discards this algorithm, if we assume admitted

connections' QoS should always be met (which seems a major reason to introduce admission control algorithm in the network). The ARAC on the other hand offers higher blocking probabilities but at the same time offers much greater flexibility when choosing averaging interval, as it's blocking and dropping probabilities remain virtually constant for different values of averaging intervals. At the same time, when using ARAC the QoS requirements of all ongoing connections were met for all simulated cases (Fig. 9).

5 Future Work

Future work will include developing MBAC algorithms that cooperate tightly with congestion control (CC) algorithms and therefore providing means of keeping blocking and dropping probabilities at predefined levels. Moreover authors plan to implement the ertPS class of service as ertPS is best suited for VoIP with silence suppression (e.g. VBR).

Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 216203 (DaVinci project).

References

1. The Network Simulator 2 project, <http://www.isi.edu/nsnam/ns/>
2. Flizikowski, A., Kozik, R., Gierszal, H., Przybyszewski, M., Hołubowicz, W.: WiMAX system level simulation platform based on ns-2 and DSP integration. In: Broadbandcom, Wrocław, Poland (2009)
3. Flizikowski, A., Przybyszewski, M., Majewski, M., Kozik, R.: Evaluation of guard channel admission control schemes for IEEE 802.16 with integrated nbD-PCcodes. In: International Conference on Ultra Modern Telecommunications, ICUMT St.Petersburg, Russia (2009)
4. Flizikowski, A., Przybyszewski, M., Majewski, M., Kozik, R.: Performance comparison of guard channel admission control schemes for IEEE 802.16 system with various turbo code FEC schemes. In: 1st IEEE International Workshop on Broadband Wireless Network Communications Performance In conjunction with IEEE IPCCC 2009, Phoenix, USA (2009)
5. Flizikowski, A., Hołubowicz, W., Przybyszewski, M., Grzegorzewski, S.: Admission control and system capacity assessment of WiMAX with ACM and nb-LDPC codes - simulation study with ViMACCS ns2 patch. In: The International Conference on Advanced Information Networking and Applications, Perth, Australia,
6. FP7 DaVinci project, <http://www.ict-davinci-codes.eu/>

7. Aymen, B., Loutfi, N.: Design and implementation of a QoS-included WiMAX module for NS-2 simulator. In: SIMUTools 2008, March 3-7, Marseille, France (2008)
8. Lakkakorpi, J., Sayenko, A.: Measurement-Based Connection Admission Control Methods for Real-Time Services in IEEE 802.16e. In: Second International Conference on Communication Theory, Reliability, and Quality of Service (2009)
9. FP7 DaVinci project, <http://www.ict-davinci-codes.eu/>
10. EuQoS. FP7 project, <http://www.euqos.eu>

The Gap between Packet Level QoS and Objective QoE Assessment of WWW on Mobile Devices

Adam Flizikowski¹, Damian Puchalski^{1,2},
Kacper Sachajdak¹, and Witold Hołubowicz³

¹ University of Technology and Life Sciences,
Bydgoszcz, Poland
e-mail: adamfli@utp.edu.pl, livetofly@wp.pl

² ITTI Ltd, Poznań, Poland
e-mail: damian.puchalski@itti.com.pl

³ Adam Mickiewicz University, Poznań, Poland
e-mail: holubowicz@amu.edu.pl

Summary. This paper studies the difference between packet level evaluation of mobile WWW and novel image based method for objective quality of experience (QoE) assessment. Authors, based on the reference web browser architecture, show the two kinds quantitative results of measurements with mobile phone and different browsers (Opera Mini and default browser). The designed test methodology proves that there is a gap in the key performance indicators values depending on the measurement method used. Image based assessment of the WWW QoE is the right choice for mobile operators when evaluating end user satisfaction.

1 Introduction

The mobile phone market since many years has been one of the most rapidly growing sectors of telecommunications. Nowadays is not easy to imagine everyday life without mobile phone existence and support. In the last years mobile Internet access has reached popularity close to the legacy service of the voice calls and text messages (SMS). Over the past years mobile terminal evolved from simple voice communication device to technologically advanced device which provides multiple capabilities in distinct areas of entertainment and the mobile Internet. Popularity of cell phones in comparison to other portable devices is shown in results of Hagggle Project [1]. These results proved that mobile phone and laptop devices are the most popular among other portable devices. According to the IP traffic forecast in mobile devices provided recently by Cisco [2], the IP traffic will raise 66-fold between years 2008 and 2013. The same studies predict that in the year 2013 over 80% of data exchange will be transmitted utilizing mobile terminals and laptops

with wireless connectivity at the level higher than throughputs in present 3G standard. Relatively small part of overall IP traffic will be generated by stationary devices using 4G networks (e.g. WiMAX). It can be seen that laptops, PDAs and handsets (mainly smartphones which will evolve to the next generation of devices) will be much more popular.

The trend of growing monthly transfer generated by single mobile shows that the maximum available data volume per mobile packet networks in range of 1 and 5GB will be limiting factor in the coming years, even for users who use mobile data communication for performing small to medium number of simple tasks requiring the network. Rapid growth of the new kind of services implemented at the network premises (e.g. cloud computing) and continuous evolution of mobile devices capabilities mean that the single terminal gives much more possibilities than ever if provided the Internet access. Expansion of mobile access to data networks and its spectacular success will follow.

The structure of the paper is as follows: Section 2 presents related work in the area of the QoE evaluation for WWW services, current market analysis is the subject of Section 3. Appropriate standards and regulations for measuring QoE/QoS of telecommunication services are described in Section 4. Original methodology designed by authors to measure the gap between packet based QoS and user-centric QoE is introduced in Section 5. Results of the measurements along with conclusions are provided in Sections 6 and 7 accordingly.

2 Related Work

The increasing amount of non-voice traffic on the networks calls for the need to provide dedicated methods to determine the level of end-user acceptance. Current literature widely describes aspects of QoE measurements related with the current work of the authors but main differences are easy to be identified.

The research by J.Nielsen [3] presents a survey-based assessment of the usability as perceived by the end-user of mobile website rendered using mobile phone. The key difference between these tests and the work by the authors is the subjectivity of the method used. The aim of Coverdale [4] is to show a method of estimating the end to end performance in IP networks. The estimation is based on calculations of time transfer in an every phase of connection. Another paper presents a system which was developed to evaluate QoE on IP networks [5]. The system's architecture is designed to be capable of emulating multi-agent networks along with dynamically changing conditions. The experiment was conducted on the basis of ITU-T Recommendation G.1030 (namely Annex A). Assessment was prepared using legacy PCs, not mobile phones, so it differs with the tests presented later in this paper, but the main idea of the approach is the use of the timeline of connection and timings between particular actions. Similar approach was presented in article [6]. In the paper [7] authors analyze the impact of router queuing disciplines on

the multimedia QoE performance in IPTV. The analysis is based on advanced performance measurements of objective QoE metrics in over 500 test scenarios in an IPTV testbed. Similarly this approach focuses on measuring objective QoE but the subject of assessments is different.

3 Market Analysis

The previous sections show general trends and forecasts of future use of the mobile phone, this section is focused on mobile technologies market analysis. The continuous technological evolution of devices is a natural step needed to cope with growing users expectations. The continuous improvement in hardware performance and rise of new technologies facilitate the increase of total mobile telephony users, which in 2007 in Poland has achieved level of 38 millions. Moreover forecasts claim it will reach 45 millions in year 2012 [8]. The comparison of popularity of cell phone models used by polish Internet users can be found in [9]. Ranking shows the major changes took place on the market of mobile devices in the last 2-3 years. According to this research, only two models (Nokia N95 and iPhone) stayed at the lead on the mobile phone market, while the rest underwent a high rotation of the models. This means that a typical user of mobile Internet has access to a device with greater possibilities in area of WWW services each year. The more advanced terminals are displacing legacy ones. Also this trend is fostering today's success of the mobile Internet.

4 Regulations and Standards (QoE/QoS) in Mobile Networks

Great popularity of WWW services (measured by the number of hits in the biggest portals in Poland is close to several billions per month [8]) browsed using mobile devices forces operators to guarantee suitable service quality to end-user. The end-user is the most important part of mobile market, so it is very desirable to undertake actions which will cause increase of the quality of service level, and interest in mobile Internet. These WWW performance evaluation actions may be performed from different points of view:

- Quality of Service (QoS) - defined as set of network indicators characterizing telecommunication services, constituting its quality and stating their usefulness to satisfy user expectations. QoS is focused on absolute indicators, specifying technical features of service,
- Quality of Experience (QoE) - defined as level of user satisfaction of services provided by operator. The measurement of QoE level allows to (in subjective or objective way) determine "is a user satisfied with the service?". QoE indicators are used to designate the minimum level of quality which can satisfy the user.

Authors of many publications propose QoE models that support measurement of network performance with users satisfaction of using services, but none of those solutions meet demands and recommendations from ITU-T G.1000 series. The existing solutions dedicated to measure QoE level can be grouped into two main categories depending on the availability of the reference signal [9]:

- black-box - based on analyzing system under test by comparing the received and reference multimedia signal,
- glass-box - in which results of measurements are various key performance indicators (KPI) of a service based on the observed intra-system (pocket) level parameters (e.g. jitter, delay).

In turn methods for QoE measurement belonging to the black-box group are divided into four subcategories [10]:

- PESQ (Perceptual Evaluation of Speech Quality) - described in ITU-T P.862 standard based on psychoacoustic attributes of human hearing,
- PEAQ (Perceptual Evaluation of Audio Quality) - described in ITU-R BS.1387 standard based on psychoacoustic attributes of human hearing,
- PEVQ (Perceptual Evaluation of Video Quality) - based on psychovisual attributes of human sight (standardization procedure in progress),
- PEDQ (Perceptual Evaluation of Data-link Quality) - described in the ITU-T G.CHIRP standard.

The document which is regulating norms and specifications corresponding to networks and electronic communications services in European Union Members is Commission Decision released 11th of December 2006 [11]. More of those regulations are norms developed by European Telecommunications Standards Institute (ETSI) - ETSI EG 202 057 (chapters 1-4) [12] and 202 009 (chapters 1-3) [13]. Authors of the standards form the perspective of measurement methods recommend test connections, continuous monitoring of network traffic or combinations of both of them. Those methods are focusing on the QoS measurement at the packet level just as adopted by polish Telecommunication Law. The indicators contained there form the unified catalogue of indicators, together with definitions, measurement methodologies and methods of data processing and results analysis [11]. Consequently the methods of measuring user satisfaction when using mobile WWW services are ignored, both in EU regulations and polish Telecommunication Law. Authors of current paper will show in the following sections that measuring only the network factors may result in a very inadequate level of user satisfaction.

5 Methodology

The difference between QoE and QoS with respect to measuring WWW services using mobile browsers is well illustrated on reference model of a browser, described widely in [10]. Based on this article authors assume that

when using mobile browser many factors influence end-user satisfaction. These factors depend not only on a mobile network but also on a browser (interface, rendering and browsing engine, etc.). Putting things straightforward - it is not guaranteed that the good packet level QoS statistics of WWW service will assure high level of end user satisfaction. Authors conducted an experiment which was intended to show differences between results of QoS and QoE approach. For this case authors have performed:

- measurement of downloaded data in a QoE approach via GSM/UMTS network. The reference testing pages - prepared to obtain differences in the volume of downloaded data between the mobile phone built-in counters and the actual size,
- measurement of downloaded data in a QoS approach, where access to the reference websites was performed by WLAN - in order to compare results with results obtained in measurement described above,
- measurement of WWW page downloading time performed by simultaneously running the two cases above. The aim was to compare obtained results and to find the answer which of the methods is more adequate to assess a user satisfaction.

In the following sector we will describe only the last of the three above mentioned experiments, due to the fact that it provides best answer to the question whether the QoS approach for mobile Internet quality measurement is appropriate. To perform the test authors used the Nokia E66 mobile phone with Opera Mini browser (version 5.0). Three reference web pages with different content type (only text, only pictures and the combination of both of them) and size of 200kB were used. Testbed architecture for performing measurements is shown below (Fig. 11).

In the proposed architecture we have used Squid proxy tool to measure the throughput and volume of downloaded web pages in the downlink direction. It is installed just before the WLAN access point. In addition an external video camera was attached to the measurement terminal and recorded the process of repeated (10-times) downloading of each of the three test web pages from the mobile phone. The above instrumentation allows the subject of recordings to be simultaneously both the phone's display and the laptop's screen with the tool Squid running (with packet statistics visible). The video recordings aim to determine absolute time difference (gap) between arrival of the last packet in the network and the moment of web page download completion (exactly as viewed by the user). This approach enabled the comparison of the overall web page downloading time from the perspective of both evaluated approaches (QoS/QoE). In the measurements the granularity of 40ms was obtained, resulting from the number of 25 video frames per second of recording.

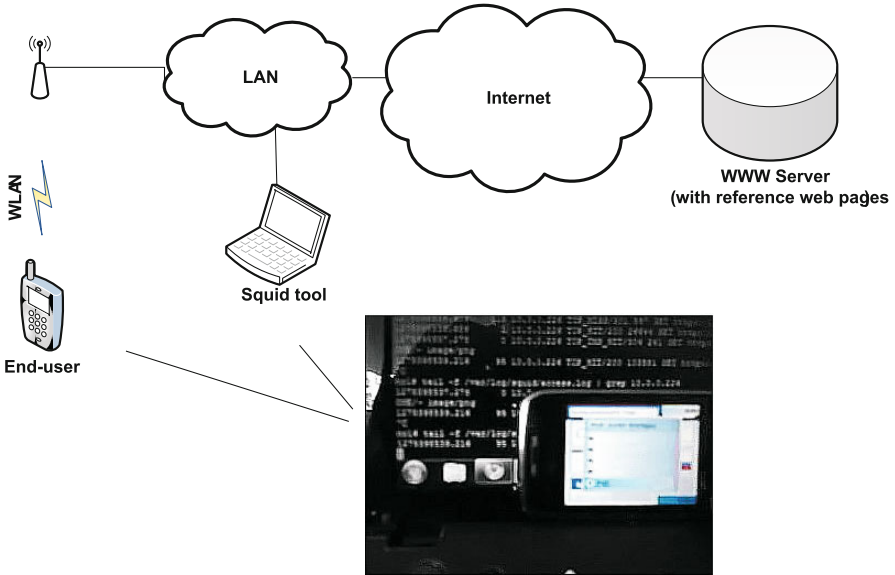


Fig. 1. Conceptual architecture of the measurement components

6 Results

The averaged page download measurement results were collected in the table (Table 1).

Table 1. Averaged measurement results

Content	TEXT			TEXT + IMAGE			IMAGE		
	Download time [s]	Last packet	shift [s]	Download time [s]	Last packet	shift [s]	Download time [s]	Last packet	shift [s]
---	Squid tool	Phone Screen	shift [s]	Squid tool	Phone Screen	shift [s]	Squid tool	Phone Screen	shift [s]
Average [s]	3,5	7,8	0,9	2,3	6,6	0,9	1,0	5,9	1,0
Standard Deviation [s]	0,38	0,04	0,01	0,30	0,04	0,10	0,05	0,32	0,06

Graphical interpretation of these results is presented in figure below (Fig. 2).

At the time t_0 (beginning of the graph) the OK button was pressed on the phone. From the user perspective this is the beginning of page download process. User starts to wait for a loaded content to fully download. The red

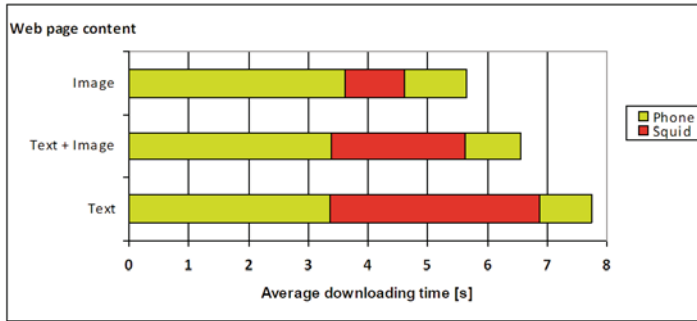


Fig. 2. Comparison of average download time measured by Squid tool and the phone's screen

highlighted bars in the figure (Fig. 2) indicate the time when Squid tool has detected the transfer of packets in the network level. As shown in the graph the first incoming packet in the network is about 3-4 seconds after pressing the button which started the download process. The last packet is coming about a second before the completion of web page download (indicated by the progress bar disappearance). Time intervals (Fig. 2) vary significantly for the web pages with different content. Interpretation of this graph is as follows:

- green bar (with the invisible part hidden under the red bar) - represents the approach with QoE assessment - this is the page load time from the perspective of the end-user, as seen by him on the screen,
- red bar - represents results of the QoS approach based on the measurement of network statistics - in this case red box describes the time elapsed from the appearance of the first to in the last packet at the downloading session.

7 Conclusions

The experiments described in previous sections prove the existence of significant differences in web page download performance when different approaches to quality measurement are applied. The amount of downloaded data may play a minor role in user satisfaction assessment (due to the existence of mobile versions of regular web pages, browsers offering higher compression standards or operator offering flat rate for using mobile WWW) but the time of opening even a simple (without graphics) page can be main problem that causes users dissatisfaction. According to the QoS approach test pages were opening in about 2.5sec which at the size of 200kB gives a throughput on the level of 80kB/s. However the overall time of connection, processing and content rendering of the same page can be around 6.5sec, which gives a

throughput about 30kB/s, and thus the differences is very significant (from the one with QoE focus). This assessment clearly confirms that providing a suitable level of quality at the level assessed by the QoS approach may fail to provide a satisfactory level of mobile WWW services on a mobile phone.

References

1. Huggle.: An Innovative Paradigm for Autonomic Opportunistic Communication. Deliverable 6.1: User-Centred Design of Huggle Applications (2006)
2. Cisco Systems Inc. The Cisco Visual Networking Index (VNI) Global Mobile Data Traffic Forecast (2009)
3. Nielsen, J.: Alterbox, Mobile Usability (2009)
4. Coverdale, P.: G1030 - Estimating End-to-end Performance in IP Networks. Nortel Networks, Canada (2003)
5. Ibarrola, E., Iberal, F., Taboada, I., Ortega, R.: Web QoE Evaluation in Multi-Agent Networks. Department of Electronics and Telecommunications University of the Basque Country. In: UPV/EHU and Department of Industry, System Integration Northline, Atos Origin S.A.E, Bilbao (2009)
6. Shaikh, J., Fiedler, M., Collange, D.: Quality of Experience From User and Network perspectives (2009)
7. Calyam, P., Chandrasekaran, P., Trueb, G., Howes, N., Yu, D., Liu, Y., Xiong, L., Ramnath, R., Yang, D.: Impact of Router Queuing Disciplines on Multimedia QoE in IPTV Deployments (2009)
8. Gemius, Raport Polski Internet 2008/2009 (2009)
9. <http://www ranking.pl/pl/rankings/mobile-devices-models>
10. Flizikowski, A., Choraś, M., Wachowiak, M., Hołubowicz, W.: A Method For Semi-automated Assessment of User Satisfaction When Using WWW Services with Mobile Terminals. In: International Conference on Image Processing and Communications, Bydgoszcz (2009)
11. Dziennik Urzędowy Unii Europejskiej, Decyzja Komisji z dn. 11 XII, Ustanawiająca Wykaz Norm i Specyfikacji Dotyczących Sieci i Usług Łączności Elektronicznej (2006)
12. ETSI EG 202 057: Speech Processing, Transmission and Quality Aspects (STQ), User Related QoS Parameter Definitions and Measurements (2005)
13. ETSI EG 202 009: User Group, Quality of Telecom Services (2007)
14. UKE, Zadanie SP II.9 Opracowywanie i Wdrażanie Nowych Metod Pomiarowych. Etap I - Katalog Wskaźników Jakości Różnych Usług wraz z ich Definicjami (2007)

Evaluation of Smoothing Algorithms for a RSSI-Based Device-Free Passive Localisation

Gabriel Deak, Kevin Curran, and Joan Condell

Intelligent System Research Centre,
Northland Road, Londonderry, UK
e-mail: {Deak-G, K.J.Curran, J.Condell}@ulster.ac.uk

Summary. There are a number of techniques used in modern Location aware systems such as Received Signal Strength Indicator (RSSI), Time of Arrival (TOA), Time Difference of Arrival (TDOA) and Angle of Arrival (AOA). However the benefit of RSSI-based location positioning technologies, is the possibility to develop location estimation systems without the need for specialised hardware.

The human body contains more than 70% water which is causing changes in the RSSI measurements. It is known that the resonance frequency of the water is 2.4 GHz. Thus a human presence in an indoor environment attenuates the wireless signal. Device-free Passive (DfP) localisation is a technique to detect a person without the need for any physical devices i.e. tags or sensors. A DfP Localisation system uses the Received Signal Strength Indicator (RSSI) for monitoring and tracking changes in a Wireless Network infrastructure. The changes in the signal along with prior fingerprinting of a physical location allow identification of a person's location. This research is focused on implementing DfP Localisation built using a Wireless Sensor Network (WSN). The aim of this paper is the evaluation of various smoothing algorithms for the RSSI recorded in a Device-free Passive (DfP) Localisation scenario in order to find an algorithm that generates the best output. The best output is referred to here as results that can help us decide if a person entered the monitored environment. The DfP scenario considered in this paper is based on monitoring the changes in the wireless communications due to the presence of a human body in the environment. Thus to have a clear image of the changes caused by human presence indoors, the wireless recordings need to be smoothed. We show results using algorithms such as five-point Triangular Smoothing Algorithm, 1-D median filter, Savitzky-Golay filter, and Kalman filter.

1 Introduction

The possibility of estimating a position represents a crucial component currently. Estimating the position can be very useful for many applications such as determining the location of assets, monitoring patients in a hospital,

security, tour guides, conference guides, shopping guides and information, network access based on the user's location, and games [1, 2]. The Global Positioning System (GPS) was first used for military applications [3]. GPS is a popular technology worldwide for outdoor location estimation. The number of devices which use a GPS is increasing every year due to the current low cost of implementing GPS technology. However GPS does not work indoors, in cities with tall buildings, or between mountains due to the requirement of Line-of-Sight (LoS) for the communication with the satellites. Thus, there is a need for accurate indoor location estimation technologies [4].

Location estimation technologies can be separated into indoor and outdoor localisation systems. Various approaches such as location fingerprinting (scene analysis), triangulation, trilateration, hyperbolic lateration, proximity, and dead reckoning [1, 3, 5, 6] were used in location sensing systems. The metrics used frequently for developing tracking systems are: Received Signal Strength Indicator (RSSI), Time of Arrival (TOA) / Time Difference of Arrival (TDOA), and Angle of Arrival (AOA) or Direction of Arrival (DOA). Possible technologies for indoor systems are: WLANs, Through Wall Imaging (TWI) using Ultra-wideband (UWB), Field strength systems, Radio-frequency Identification (RFID), and Next-generation indoor positioning systems [7].

Location tracking techniques can be classified into two categories: 1) systems requiring tracked persons to participate actively; and 2) systems using passive localisation. The two classes are also known as active and passive tracking systems. By participating actively we refer to the tracked person as carrying an electronic device sending information to the system, information used to infer the user's position. In the passive localisation case the position is estimated based on the variance of a measured signal or video processing.

Many of the proposed technologies use the fingerprinting method to estimate the location of a tracked person. Fingerprinting in location estimation systems refers to a method that compares the *fingerprint* of some characteristic of a signal that is location dependent. This algorithm has two stages: the offline stage and the online stage [1, 6, 8, 9, 10]. The techniques that employ database comparison are fingerprinting, pattern recognition, or pattern matching [5, 11]. The fingerprint method is presented in [12] as a *map* of the coverage area of the network obtained by measuring received signal strength indicator (RSSI).

The AT&T Cambridge's Active Bats system [13] uses physical devices such as ultrasonic badges. Ubisense and Ekahau are commercial location estimation systems. Ubisense has a very high precision approximately 15 cm due to the use of the active tags signal triangulation [14, 15]. The Ekahau systems, RADAR systems [1], and LA200 systems [16] use WLAN fingerprinting which is considered to be the most successful method. MIT Cricket system use tags incorporating RF radio chip, a microcontroller, and an ultrasonic receiver [17].

The RSSI measurements used to build a fingerprint are strongly dependent on the line-of-sight (LOS) between the sending device and the receiving device. Therefore in most of the systems which use fingerprint methods, RSSI is recorded for different orientations (north, south, west, east or 0° , 90° , 180° , 270°) at each point in the environment. RSSI-based position estimation techniques can be classified as terminal assisted, terminal based, and network based [18]. The terminal assisted mode is based on RSSI measurements taken by the target and sent to a server which is managing the radio map and is also running the algorithm to compute the target's location. In the terminal based mode the radio map is built on the terminal and used to determine the target's position. The network based method uses RSSI measurements taken by the Access Points (APs) or Base Stations (BSs) listening to the beacons in the environment.

2 Initial Measurements

RSSI measurements are very noisy, thus the signal amplitude representations are not smooth. The aim of smoothing the recorded data is to obtain a signal which can clearly show the variance of the signal when a person is walking or standing causing interference in the wireless communication. Having data with a lower noise level can be helpful for setting up a threshold. When the signal exceeds this threshold, an event can be triggered showing the presence of a person in the indoor environment.

Tests were conducted and the results are shown in Fig. 1.

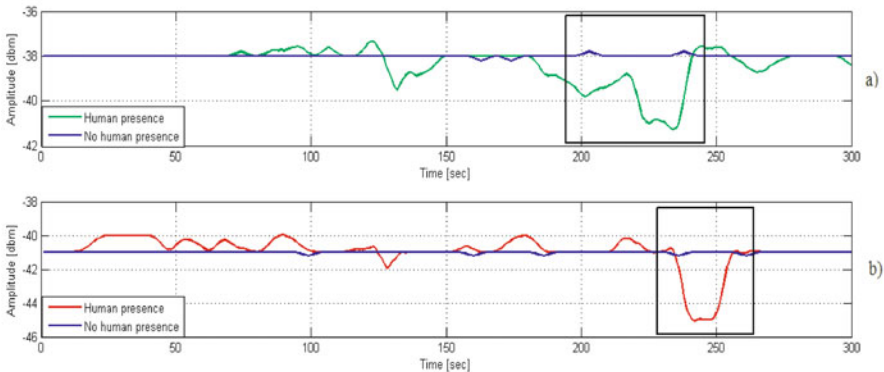


Fig. 1. Experimental results with two Sunspots notes using a five-point Triangular Smoothing Algorithm: a) first node; b) second node

The recorded RSSI was smoothed with a five-point triangular smoothing algorithm using the *fastsmooth* matlab function:

$$s = \text{fastsmooth}(\text{data}, \text{width}, \text{type}, \text{edge}) \quad (1)$$

where s is a vector which returns the smoothed data, data represents the vector containing the RSSI measurements, smooth width is a positive integer and type refers to the type of the smoothing algorithm. In our case $\text{width}=5$ and $\text{type}=2$ for a five-point smoothing algorithm. If $\text{type}=1$ than we have a rectangular smoothing algorithm and for $\text{type}=3$ a pseudo-Gaussian smoothing algorithm.

In Fig. 2, a median filter was used to filter the noise in the signal. The filter was applied on the same data as the previous filter. The *medfilt1* function from Matlab was used to obtain the results presented in Fig. 2.

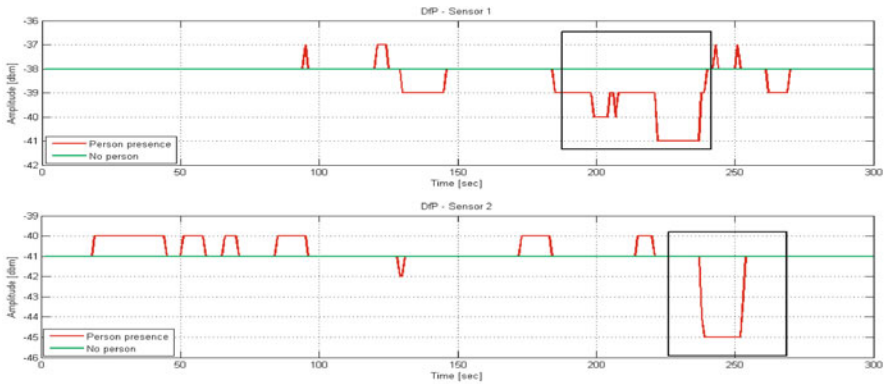


Fig. 2. 1-D median filtering

The syntax of the 1-D median filter is given by Equation 2:

$$y = \text{medfilt1}(x, n) \quad (2)$$

where x is the vector containing our RSSI data. If n is odd $y(k)$ is the median of $(x(k-(n-1)/2), x(k+(n-1)/2))$ and if n is even $y(k)$ is the median $(x(k-n/2), x(k-(n/2)+1), \dots, x(k+(n/2)-1))$.

Fig. 3 present the results using a Savitzky-Golay smoothing filter which is also called a digital smoothing polynomial filter or a least-squares filter.

The Savitzky-Golay filter usually performs better than the standard averaging filters which tend to filter out signal's high frequency together with the noise. Although the Savitzky-Golay filters are effective preserving the signal's high frequency, they are not as successful as standard averaging filters at rejecting noise. The graphics were plotted using Matlab's function for the Savitzky-Golay filter with the following syntax:

$$y = \text{sgolayfilt}(x, k, f) \quad (3)$$

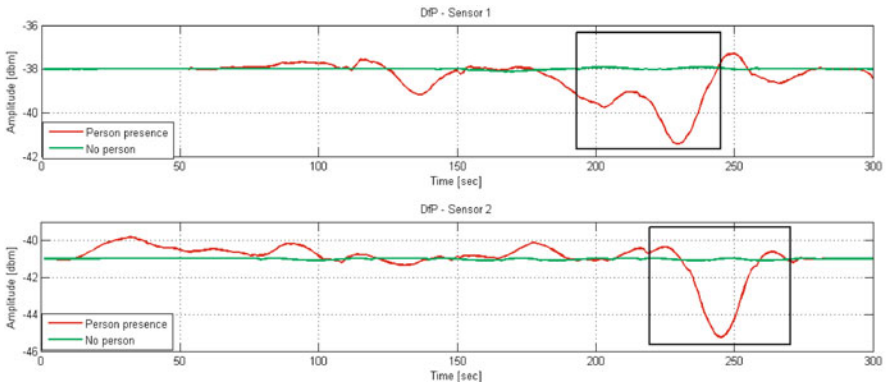


Fig. 3. Savitzky-Golay Finite Impulse Response (FIR) smoothing filter

The Kalman filter [3] was also used for smoothing the data recorded from one of the free-range node. The results in Fig. 4 are plotted using an implementation of the Kalman filter in Matlab.

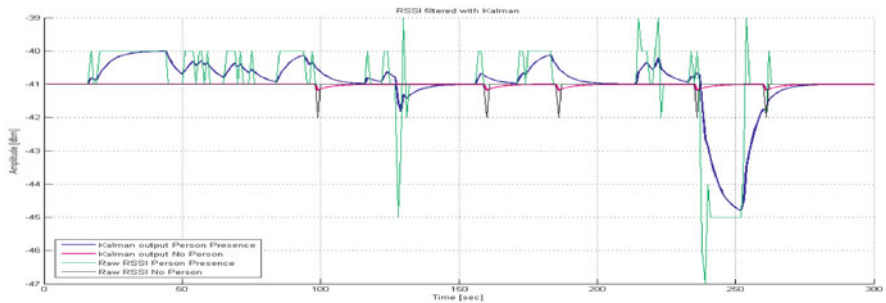


Fig. 4. Smoothing RSSI data using Kalman Filter

The Kalman filter is known to be used for noisy measurements and has two distinct sets of equations: Time update (prediction) and Measurement update (correction). Future work will investigate the use of other smoothing algorithms. We will continue to focus on finding the algorithm which filters the RSSI data and offers us the best output for choosing a threshold necessary to trigger an event when the presence of a person is detected.

3 Evaluation

The test bed in Fig. 5 (a) consists of two free-range Sunspot motes with fixed positions sending beacons every 500 msec. The messages sent by each node contain the following parameters: RSSI, node ID, battery voltage and battery

level. The battery voltage and the battery level are being measured due to the fact that wireless communication could be affected if the node's battery is running low.

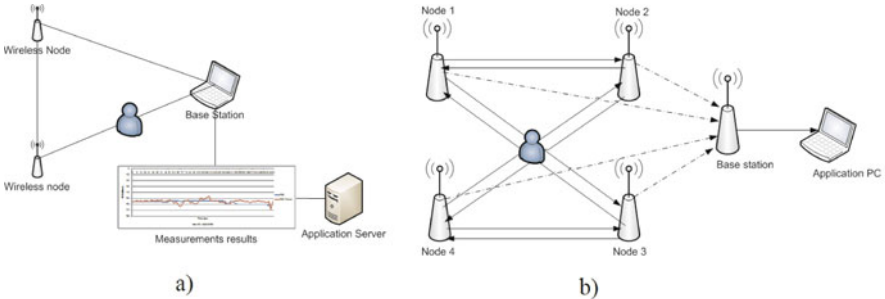


Fig. 5. Experiments: a) current test bed; b) future test bed

The nodes are placed 1.5m away from the base station with a distance of 1.5m separating the nodes and a 1m height. Fig. 5 (b) shows the future test bed. Current work involves investigation of the firmware that will run on the four free-range Sunspot nodes. In order to implement an application using WSNs, it is necessary to take into account the design and resources constraints which are specific for WSN nodes. The protocols need to be carefully designed in order to preserve as much energy as possible putting the nodes into sleep mode whenever they are not used. It is also important to consider WSNs specific limitations such as short communication range, bandwidth, processing power and storage space.

The free-range Java Sunspot nodes are built from four components: 1) the sunroof; 2) a sensor board including a 2G/6G three-axis accelerometer, a temperature sensor, eight LEDs, a light sensor, two programmable switches, I/O pins, and four output pins; 3) a processor board running a 180 MHz 32 bit ARM920T core - 512K RAM/4M Flash; 4) an internal 3.7V 720mAh lithium-ion rechargeable battery.

4 Conclusion

Four types of smoothing algorithms were used for filtering the data. Examples of more complex methods which could be considered in the future as smoothing algorithms are Hidden Markov Models (HMM) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) architectures. The average error was computed for each algorithm presented and the results showed that the 1-D median filtering has the smallest average error compared with the rest of the filters due to the fact that small variations of the signal are considered to be zero. The filters can be optimised further by modifying or adding more parameters.

Acknowledgements

This work is supported by a Vice Chancellor Research Scholarship from the University of Ulster. We would also like to thank Oracle for their help in supplying Java Sunspot Development Kits.

References

1. Kolodziej, K.W., Hjelm, J.: Local positioning systems: LBS applications and services. CRC Press, Boca Raton (2006)
2. Zhou, R.: Wireless indoor tracking system (WITS). In: Aktuelle Trends in der Soft-wareforschung, Tagungsband zum doIT Software-Forschungstag, dpunkt, Heidelberg, Germany, pp. 163–177 (2006) (to appear)
3. Krumm, J.: Ubiquitous Computing Fundamentals. CRC Press, Boca Raton (2010)
4. Kupper, A.: Location-Based Services: Fundamentals and Operation. Wiley, Chichester (2005)
5. Bensky, A.: Wireless positioning technologies and applications. Artech House, Boston (2007)
6. Liu, H., Darabi, H., Banerjee, P., Liu, J.: Survey of Wireless Indoor Positioning Techniques and Systems. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)* 37(6), 1067–1080 (2007)
7. Yu, K., Sharp, I., Guo, Y.J.: Ground-Based Wireless Positioning. Wiley-IEEE Press (2009)
8. Ito, S., Kawaguchi, N.: Bayesian Based Location Estimation System Using Wireless LAN. In: Third IEEE International Conference on Pervasive Computing and Communications Workshops, pp. 273–278 (2005)
9. Nerguizian, C., Despins, C., Affes, S.: Indoor geolocation with received signal strength ingerprinting technique and neural networks. *LNCS*, pp. 866–875 (2004)
10. Stoyanova, T., Kerasiotis, F., Prayati, A., Papadopoulos, G.: Evaluation of impact factors on RSS accuracy for localization and tracking applications in sensor networks. *Telecommunication Systems* 42(3-4), 235–248 (2009)
11. Curran, K., Furey, E.: Pinpointing users with location estimation techniques and Wi-Fi hotspot technology. *International Journal of Network Management* 16(5) (2007)
12. Munoz, D., Lara, F.B., Vargas, C., Enriquez-Caldera, R.: Position Location Techniques and Applications. Academic Press, London (2009)
13. Addlesee, M., Curwen, R., Hodges, S., Newman, J., Steggles, P., Ward, A., Hopper, A.: Implementing a sentient computing system. *Computer*, 50–56 (2001)
14. Steggles, P., Gschwind, S.: The Ubisense smart space platform. In: Adjunct Proceedings of the Third International Conference on Pervasive Computing, vol. 191, pp. 73–76 (2005)
15. Ubisense. Ubisense System (2010)

16. Griffith, E.: *New Finds in Real-Time Location* (2007)
17. Priyantha, N.B., Chakraborty, A., Balakrishnan, H.: The cricket locationsupport system. In: *Proceedings of the 6th annual international conference on Mobile computing and networking*, pp. 32–43. ACM, New York (2000)
18. Curran, K., Norrby, S.: RFID-Enabled Location Determination within Indoor Environments. *International Journal of Ambient Computing and Intelligence* 1(4), 63–86 (2009)

Performance Evaluation of ADS System Based on Redundant Dictionary

Rafał Renk^{1,3}, Łukasz Saganowski²,
Michał Choraś^{1,2}, and Witold Hołubowicz^{1,3}

¹ ITTI Ltd., Poznań

e-mail: rafal.renk@itti.com.pl

² Institute of Telecommunications,

University of Technology and Life Sciences, Bydgoszcz

e-mail: chorasm@utp.edu.pl

³ Adam Mickiewicz University, Poznań

e-mail: renk@amu.edu.pl

Summary. Anomaly detection approach is a new, emerging trend for network security especially for high-security networks (such as military or critical infrastructure monitoring networks). In our previous work we proposed a new methodology for such intrusion detection systems. We proposed new signal based algorithm for intrusion detection on the basis of the Matching Pursuit algorithm. As to our best knowledge, we are the first to use Matching Pursuit for intrusion and anomaly detection in computer networks. Hereby, we report further, more extensive, evaluation of the proposed methodology. We show results for 15 metrics characterizing network traffic (previously we tested our system using packets-per-second only). Moreover, we provided the comparison of our method with state-of-the-art DWT-based anomaly detection system and proved that our solution gives better results in terms of detection rate and false positives.

1 Introduction

Most current IDS systems have problems in recognizing new attacks (0-day exploits) since they are based on the signature-based approach. In such mode, when system does not have an attack signature in database, such attack is not detected. Another drawback of current IDS systems is that the used parameters and features do not contain all the necessary information about traffic and events in the network [1].

In our previous work (e.g. [2]) we proposed a new solution for ADS (Anomaly Detection System) system based on signal processing algorithm. ADS analyzes traffic from internet connection in certain point of a computer network. The proposed ADS system uses redundant signal decomposition method based on Matching Pursuit algorithm. Matching Pursuit is a known

signal processing technique used for instance in audio compression, image and video compression. However, as to our best knowledge, we are the first to use Matching Pursuit for intrusion and anomaly detection in computer networks.

ADS based on Matching Pursuit uses Dictionary of Base Functions - *BFD* to decompose input 1D traffic signal (1D signal may represent for example packets per second) into set of based functions called also atoms. The proposed *BFD* has ability to approximate traffic signal.

In this paper we present further evaluation of our methodology. Previously, we reported the results only for one network traffic parameter (packets per second). Hereby, we show results for 15 network traffic parameters (from the benchmark DARPA database). Moreover, we compare our results to the method based on Discrete Wavelet Transform and prove that we achieve better results.

2 Anomaly Detection Algorithm Based on Redundant Dictionary of Base Functions

In this section we shortly present our implementation of the ADS system based on Matching Pursuit [3, 4, 5].

In the basic Matching Pursuit algorithm atoms are selected in every step from entire dictionary which has flat structure. In this case algorithm causes significant processor burden. Therefore, a dictionary with internal structure was used in our coder. Such dictionary is built from: atoms and centered atoms. Centered atoms group such atoms from D that are as correlated (to each other) as possible. To calculate the measure of correlation between atoms, the function $o(a, b)$ can be used [6]:

$$o(a, b) = \sqrt{1 - \left(\frac{|\langle a, b \rangle|}{\|a\|_2 \|b\|_2} \right)^2} \quad (1)$$

The quality of centered atom can be estimated according to (2):

$$O_{k,l} = \frac{1}{|LP_{k,l}|} \sum_{i \in LP_{k,l}} o(A_{c(i)}, W_{c(k,l)}) \quad (2)$$

$LP_{k,l}$ is a set of atoms grouped by centered atom. $O_{k,l}$ is mean of local distances from centered atom $W_{c(k,l)}$ to the atoms $A_{c(i)}$ which are strongly correlated with $A_{c(i)}$.

Example dictionary structure was presented in Fig. 1. Atom tree consist of root node W_g and every centroid consist of two children. Parameter A_c represents leaf nodes (without children).

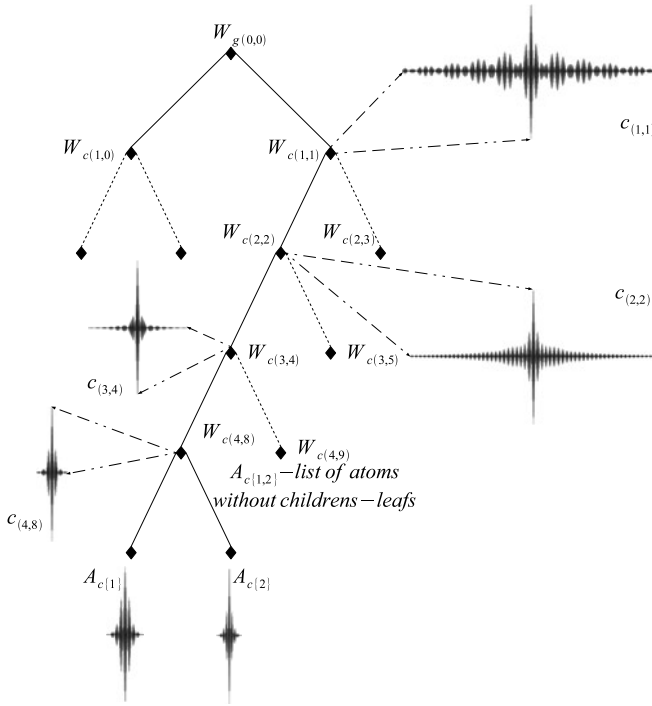


Fig. 1. Dictionary structure

Centroid $W_{c(k,l)}$ represents atoms $A_{c(i)}$ which belong to the set $i \in LP_{k,l}$. List of atoms $LP_{k,l}$ should be selected according to the Equation 3

$$\max_{i \in LP_{k,l}} o(A_{c(i)}, W_{c(k,l)}) \leq \min_{t \in D \setminus LP_{k,l}} o(A_{c(t)}, W_{c(k,l)}) \tag{3}$$

In the proposed ADS solution 1D real Gabor base function (Equation 4) was used to build dictionary [6, 7].

$$\alpha_{u,s,\xi,\phi}(t) = c_{u,s,\xi,\phi} \alpha\left(\frac{t-u}{s}\right) \cos(2\pi\xi(t-u) + \phi) \tag{4}$$

where:

$$\alpha(t) = \frac{1}{\sqrt{s}} e^{-\pi t^2} \tag{5}$$

$c_{u,s,\xi,\phi}$ - is a normalizing constant used to achieve atom unit energy.

In order to create overcomplete set of 1D base functions dictionary D was built by varying subsequent atom parameters: Frequency ξ and Phase ϕ ,

Position u , Scale s . Base functions dictionary D was created with using 10 different scales (dyadic scales) and 50 different frequencies.

Traffic Parameters are used to create one dimensional signal. This signal is decomposed with the use of Matching Pursuit transformation. After MP decomposition we achieved projection coefficients c_k which are used for creating normal traffic profiles.

Normal traffic profiles are calculated using input traffic without attack and anomalies. Normal profiles are calculated separately for every traffic feature. Our ADS system compares current traffic traces (to be analyzed) with the reference profiles calculated during normal work (stored in a database).

ADS system makes an alarm when difference between profiles exceed certain threshold (we usually use the threshold value equal to 30%).

3 Evaluation of the Proposed ADS Methodology

Performance of our approach was evaluated with the use of four trace bases [8, 9, 10, 11]. The test data contains attacks that fall into four main categories [12] such as:

1. DOS/DDOS: denial-of-service, e.g. syn flood,
2. R2L: unauthorized access from a remote machine, e.g. guessing password,
3. U2R: unauthorized access to local superuser (root) privileges, e.g., various "buffer overflow" attacks,
4. PROBING: surveillance and other probing, e.g., port scanning.

For experiments we chose 20 minutes analysis window because most of attacks (about 85%) ends within this time period [13]. We extracted 16 traffic features (see Table 1- 5) in order to create 1D signals for Matching Pursuit - Mean Projection analysis.

Traffic features were calculated with the use of 1 minute time period (Table 1 - Table 2). Traffic feature called *flow* requires more explanation since there are various definitions of this metric. A flow consist of a group of packets going from a specific source to a specific destination over a time period [14]. Flows are always considered as sessions between users and services. For every network traffic feature we created "normal profile" based on DARPA training base without attacks. Results achieved for testing base were compared to normal profiles in order to detect anomalies. We compared our system to other signal processing ADS system based on Wavelet transform [14].

For the same DARPA test base we achieved better anomaly detection rate. Our system is more computational complex than system based on wavelets transform but it is possible to realize it in real time.

In Table 1 and Table 2 detection rates achieved for DARPA benchmark trace base are presented. These are results achieved for two test days. Detection results were compared to the list of attacks which should exist in this two testing days.

Table 1. Detection Rate for W5D1 (Fifth Week, Day 1) DARPA [11] trace

Network Traffic Feature	Total number of attacks	Detected number of attacks	Detection Rate [%]
ICMP flows/minute	73	61	84.93
ICMP in bytes/minute	73	31	43.83
ICMP out bytes/minute	73	39	54.79
ICMP in frames/minute	73	59	82.19
ICMP out frames/minute	73	65	90.41
TCP flows/minute	73	68	94.52
TCP in bytes/minute	73	32	46.57
TCP out bytes/minute	73	31	45.20
TCP in frames/minute	73	57	79.45
TCP out frames/minute	73	54	76.71
UDP flows/minute	73	41	58.90
UDP in bytes/minute	73	52	73.97
UDP out bytes/minute	73	71	100.00
UDP in frames/minute	73	52	73.97
UDP out frames/minute	73	70	98.63

Table 2. Detection Rate for W5D5 (Fifth Week, Day 5) DARPA [11] trace

Network Traffic Feature	Total number of attacks	Detected number of attacks	Detection Rate [%]
ICMP flows/minute	68	49	72.06
ICMP in bytes/minute	68	56	82.35
ICMP out bytes/minute	68	54	79.41
ICMP in frames/minute	68	59	86.76
ICMP out frames/minute	68	56	82.35
TCP flows/minute	68	37	54.41
TCP in bytes/minute	68	41	60.29
TCP out bytes/minute	68	23	33.82
TCP in frames/minute	68	31	45.58
TCP out frames/minute	68	32	47.05
UDP flows/minute	68	66	97.05
UDP in bytes/minute	68	62	91.17
UDP out bytes/minute	68	60	88.23
UDP in frames/minute	68	62	91.18
UDP out frames/minute	68	60	88.24

In Table 3 and Table refSaganowski:idauthor:tab:4 there are results for MAWI test base. Bold numbers in tables point to existence of anomalies/attacks in certain window. In Table 5 there are results achieved for CAIDA test base. Traces consist of DDoS attacks and every trace represents 1 hour of the network traffic.

Table 3. Matching Pursuit Mean Projection - MP-MP for TCP trace (20 min. analysis window)

TCP trace (pocket/second) MAWI [8]	Window1 MPMP	Window2 MPMP	Window3 MPMP	MPMP for trace	MPMP for normal trace
Mawi 2004.03.06 tcp	210.34	172.58	239.41	245.01	240.00
Mawi 2004.03.13 tcp	280.01	214.01	215.46	236.33	240.00
Mawi 20.03.2004 tcp (attacked: Witty)	322.56	365.24	351.66	346.48	240.00
Mawi 25.03.2004 tcp (attacked: Slammer)	329.17	485.34	385.50	400.00	240.00

Table 4. Matching Pursuit Mean Projection - MP-MP for UDP trace (20 min. analysis window)

UDP trace (pocket/second) MAWI [8]	Window1 MPMP	Window2 MPMP	Window3 MPMP	MPMP for trace	MPMP for normal trace
Mawi 2004.03.06 tcp	16.06	13.80	17.11	15.65	16.94
Mawi 2004.03.13 tcp	20.28	17.04	17.40	18.24	16.94
Mawi 20.03.2004 tcp (attacked: Witty)	38.12	75.43	61.78	58.44	16.94
Mawi 25.03.2004 tcp (attacked: Slammer)	56.13	51.75	38.93	48.93	16.94

Table 5. Matching Pursuit Mean Projection - MP-MP for TCP trace with DDoS attacks (20 min. analysis window)

TCP trace (pocket/second) CAIDA [9]	Window1 MPMP	Window2 MPMP	Window3 MPMP	MPMP for trace	MPMP for normal trace
Backscatter 2008.11.15	147.64	411.78	356.65	305.35	153.66
Backscatter 2008.08.20	208.40	161.28	153.47	147.38	153.66

4 Comparison of the Matching Pursuit with Standard DWT Using 15 Traffic Parameters

In Table 6 comparison to DWT based [14] signal processing ADS was presented. Both solutions were tested with the use of DARPA [11] test traces. DARPA benchmark traces consist of attacks which belong to every layer of TCP/IP protocols stack.

In Table 6 the results for W5D1 (Week 5 Day 1) testday are reported. We used 15 traffic parameters during systems testing.

Detection Rate and false positive achieved by our methodology based on Matching Pursuit is better than in DWT based system presented in [14].

Detection rate is changing depending on the particular traffic feature. To recognize 100% of anomalies for DARPA testbed we have to use 1 to max 4 traffic features.

We also significantly reduced false positive parameter in comparison to DWT-based ADS [14]. It is very important parameter in ADS systems, since the number of false positives can not be overwhelming.

Table 6. Proposed MP-MP ADS in comparison to DWT based ADS [14]. Both solutions were tested with the use of DARPA [11] testbed (results in table are for Week5 Day1 testday; DR-Detection Rate [%], FP-False Positive [%])

Traffic Feature	MP-MP DR[%]	MP-MP FP[%]	DWT DR[%]	DWT FP[%]
ICMP flows/minute	68.49	20.54	14.00	79.33
ICMP in bytes/minute	79.45	27.39	83.33	416.00
ICMP out bytes/minute	73.97	32.87	83.33	416.00
ICMP in frames/minute	78.08	27.39	32.00	112.00
ICMP out frames/minute	72.60	30.13	32.00	112.00
TCP flows/minute	89.04	34.24	26.67	74.67
TCP in bytes/minute	47.94	32.87	8.67	23.33
TCP out bytes/minute	80.82	27.39	8.67	23.33
TCP in frames/minute	36.98	26.02	2.00	36.00
TCP out frames/minute	38.35	27.39	2.00	36.00
UDP flows/minute	89.04	41.09	10.00	74.67
UDP in bytes/minute	98.63	41.09	11.33	66.67
UDP out bytes/minute	100.00	46.57	11.33	66.67
UDP in frames/minute	98.63	39.72	12.67	66.67
UDP out frames/minute	100.00	46.57	12.67	66.67

5 Conclusion

In the article further evaluation of our methodology for Anomaly Detection Systems is presented. The major contributions of our work is a novel algorithm for detecting anomalies based on signal decomposition. In the classification/decision module we proposed to use developed matching pursuit features such as mean projection. We tested and evaluated the presented features and showed that experimental results proved the effectiveness of our method. Hereby, we reported the results achieved for 15 metrics characterizing network traffic. Moreover, comparison of our method with the state-of-the-art method based on DWT is provided, showing better results on the same DARPA traces.

Acknowledgment

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 216585 (INTERSECTION Project).

References

1. Coppolino, L., D'Antonio, L., Esposito, M., Romano, L.: Exploiting diversity and correlation to improve the performance of intrusion detection systems. In: Proc of IFIP/IEEE International Conference on Network and Service (2009)
2. Saganowski, Ł., Choraś, M., Renk, R., Hołubowicz, W.: A Novel Signal-Based Approach to Anomaly Detection in IDS Systems. In: Kolehmainen, M., et al. (eds.) ICANNGA 2009. LNCS, vol. 5495, pp. 527–536. Springer, Heidelberg (2009)
3. Mallat, S., Zhang: Matching pursuit with time-frequency dictionaries. *IEEE Transactions on Signal processing* 41(12), 3397–3415 (1993)
4. Troop, J.A.: Greed is Good: Algorithmic Results for Sparse Approximation. *IEEE Transactions on Information Theory* 50(10) (2004)
5. Gribonval, R.: Fast Matching pursuit with a Multiscale Dictionary of Gaussian Chirps. *IEEE Transactions on Signal processing* 49(5) (2001)
6. Jost, P., Vandergheynst, P., Frossard, P.: Tree-Based pursuit: Algorithm and properties. Swiss Federal Institute of Technology Lausanne (EPFL), Signal processing Institute Technical Report, TR-ITS-2005.013 (May 17, 2005)
7. Troop, J.A.: Greed is Good: Algorithmic Results for Sparse Approximation. *IEEE Transactions on Information Theory* 50(10) (2004)
8. WIDE project: MAWI Working Group Traffic Archive at, <http://tracer.csl.sony.co.jp/mawi/>
9. The CAIDA Dataset on the Witty Worm - March 19-24, Colleen Shanon and David Moore (2004), <http://www.caida.org/passive/witty>
10. <http://www.grid.unina.it/Traffic/Traces/ttraces.php>
11. Defense Advanced Research projects Agency DARpA Intrusion Detection Evaluation Data Set: <http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/index.html>
12. Lori, L., De Looze: Attack Characterization and Intrusion Detection using an Ensemble of Self-Organizing Maps. In: IEEE Workshop on Information Assurance United States Military Academy, West Point, NY, pp. 108–115 (June 2006)
13. Lakhina, A., Crovella, M., Diot, C.: Characterization of network-wide anomalies in traffic flows. In: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement, pp. 201–206 (2004)
14. Lu, W., Ghorbani, A.A.: Network Anomaly Detection Based on Wavelet Analysis. *EURASIP Journal on Advances in Signal processing*, 16 (2009), doi:10.1155/2009/837601
15. Jost, P., Vandergheynst, P., Frossard, P.: Tree-Based Pursuit: Algorithm and Properties. Swiss Federal Institute of Technology Lausanne (EPFL), Signal Processing Institute Technical Report, TR-ITS-2005.013 (May 17, 2005)
16. Kajan, E.: Information technology encyclopedia and acronyms. Springer, Heidelberg (2002)

Index

- Alhussaini Imad H., [81](#)
- Baron-Pałucka Kamila, [239](#)
- Berry C., [247](#)
- Bielecki Włodzimierz, [307](#)
- Boryna Bartosz, [373](#)
- Bosch Ramon, [213](#)
- Bujnowski Sławomir, [434](#)
- Burduk Robert, [158](#)
- Charakopoulos Stavros, [390](#)
- Cheddad Abbas, [59](#)
- Choraś Michał, [91](#), [477](#)
- Choraś Ryszard S., [121](#)
- Condell Joan, [59](#), [469](#)
- Curran Kevin, [59](#), [469](#)
- Cyganek Bogusław, [184](#)
- Czyżewski Andrzej, [150](#)
- Daunoravičienė Kristina, [255](#)
- Deak Gabriel, [469](#)
- Dembski Jerzy, [143](#)
- Demirel Hasan, [201](#)
- Dubalski Bożydar, [373](#), [400](#), [434](#)
- Eleyan Alaa, [201](#)
- Flizikowski Adam, [446](#), [460](#)
- Goumeidane Aicha Baya, [289](#)
- Griškevičius Julius, [255](#)
- Hachaj Tomasz, [271](#)
- Hamami Latifa, [289](#)
- Haraldsson Gustav Helgi, [379](#)
- Hołubowicz Witold, [446](#), [460](#), [477](#)
- Ignatovych Anatoliy, [99](#)
- Jakóbczak Dariusz, [39](#)
- Jorj Loay A., [81](#)
- Kabaciński Rafał, [104](#)
- Kadir Kurhairy Abdul, [247](#)
- Kapuściński Tomasz, [193](#)
- Kartaszyński Rafał Henryk, [230](#)
- Kiedrowski Piotr, [373](#)
- Klepaczko Artur, [262](#)
- Klimaszewski Krzysztof, [345](#)
- Korniak Janusz, [417](#)
- Korzyńska Anna, [213](#)
- Kowalski Mateusz, [104](#)
- Kozik Rafał, [91](#), [315](#)
- Krzesłowski Jakub, [175](#)
- Kulasek Łukasz, [150](#)
- Kwolek Bogdan, [336](#)
- Ledziński Damian, [434](#)
- Lejeun Marylene, [213](#)
- Lopez Carlos, [213](#)
- Mączkowski Grzegorz, [175](#)
- Majewski Jarosław, [323](#)
- Makrygiannakis Emmanouil, [390](#)
- Marchewka Adam, [353](#)
- Marciniak Tomasz, [434](#)
- Matuszak Łukasz, [360](#)
- McKevit Paul, [59](#)
- Mikołajczak Paweł, [230](#)
- Mokrzycki Wojciech, [67](#)
- Molga Agnieszka, [99](#)
- Muhsen Zahraa F., [81](#)

- Nacereddine Nafaa, [289](#)
Neuman Urszula, [213](#)
- Ogiela Marek R., [271](#)
Okarma Krzysztof, [51](#)
Olszewski Ireneusz, [425](#)
Ospina Juan, [278](#)
Oszust Mariusz, [167](#)
- Palkowski Marek, [307](#)
Panaszuk Piotr, [113](#)
Pauk Jolanta, [255](#)
Payne A., [247](#)
Pedersen Jens Myrup, [379](#), [434](#)
Pelczynski Paweł, [31](#)
Przybyszewski Marcin, [446](#)
Puchalski Damian, [460](#)
- Rajs Arkadiusz, [331](#)
Raso Jim, [255](#)
Renk Rafał, [477](#)
Riaz M. Tahir, [379](#)
Różycki Paweł, [417](#)
Rusyn Bogdan, [99](#)
Rymut Bogusław, [336](#)
- Sachajdak Kacper, [460](#)
Saeed Khalid, [113](#)
Saganowski Łukasz, [477](#)
Salamończyk Andrzej, [67](#)
- Samko Marek A., [75](#)
Sitnik Robert, [175](#)
Soraghan John J., [247](#)
Stankowski Jakub, [345](#)
Stasiak Bartłomiej, [298](#)
Strumiłło Paweł, [31](#)
Szajerman Dominik, [31](#)
Szczypiński Piotr, [31](#), [262](#)
- Śmietański Jacek, [222](#)
Śrutek Mscisław, [360](#)
- Tadeusiewicz Ryszard, [3](#)
- Varetsky Yarema, [99](#)
Vélez Mario, [278](#)
Vlissidis Andreas, [390](#)
- Wojtyna Ryszard, [323](#)
Woźniak Michał, [135](#)
Wysocki Marian, [167](#)
- Yamina Mohamed Ben Ali, [14](#)
Yogarajah Pratheepan, [59](#)
- Zabłudowski Antoni, [373](#), [400](#), [434](#)
Zabłudowski Łukasz, [400](#)
Ziou Djemel, [289](#)
Zmysłony Marcin, [135](#)