

ADVANCES IN SOFT COMPUTING 47

Ewa Pietka
Jacek Kawa (Eds.)

Information Technologies in Biomedicine

 Springer

Advances in 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

Mieczyslaw A. Kłopotek, Sławomir T. Wierzchon, Krzysztof Trojanowski (Eds.)
Intelligent Information Processing and Web Mining, 2005
ISBN 978-3-540-25056-2

Abraham Ajith, Bernard de Baets, Mario Köppen, Bertram Nickolay (Eds.)
Applied Soft Computing Technologies: The Challenge of Complexity, 2006
ISBN 978-3-540-31649-7

Mieczyslaw A. Kłopotek, Sławomir T. Wierzchon, Krzysztof Trojanowski (Eds.)
Intelligent Information Processing and Web Mining, 2006
ISBN 978-3-540-33520-7

Ashutosh Tiwari, Joshua Knowles, Erel Avineri, Keshav Dahal, Rajkumar Roy (Eds.)
Applications and Soft Computing, 2006
ISBN 978-3-540-29123-7

Bernd Reusch, (Ed.)
Computational Intelligence, Theory and Applications, 2006
ISBN 978-3-540-34780-4

Miguel López-Díaz, María ç. Gil, Przemysław Grzegorzewski, Olgierd Hryniewicz, Jonathan Lawry
Soft Methodology and Random Information Systems, 2006
ISBN 978-3-540-34776-7

Ashraf Saad, Erel Avineri, Keshav Dahal, Muhammad Sarfraz, Rajkumar Roy (Eds.)
Soft Computing in Industrial Applications, 2007
ISBN 978-3-540-70704-2

Bing-Yuan Cao (Ed.)
Fuzzy Information and Engineering, 2007
ISBN 978-3-540-71440-8

Patricia Melin, Oscar Castillo, Eduardo Gómez Ramírez, Janusz Kacprzyk, Witold Pedrycz (Eds.)
Analysis and Design of Intelligent Systems Using Soft Computing Techniques, 2007
ISBN 978-3-540-72431-5

Oscar Castillo, Patricia Melin, Oscar Montiel Ross, Roberto Sepúlveda Cruz, Witold Pedrycz, Janusz Kacprzyk (Eds.)
Theoretical Advances and Applications of Fuzzy Logic and Soft Computing, 2007
ISBN 978-3-540-72433-9

Katarzyna M. Węgrzyn-Wolska, Piotr S. Szczepaniak (Eds.)
Advances in Intelligent Web Mastering, 2007
ISBN 978-3-540-72574-9

Emilio Corchado, Juan M. Corchado, Ajith Abraham (Eds.)
Innovations in Hybrid Intelligent Systems, 2007
ISBN 978-3-540-74971-4

Marek Kurzynski, Edward Puchala, Michal Wozniak, Andrzej Zolnierok (Eds.)
Computer Recognition Systems 2, 2007
ISBN 978-3-540-75174-8

Van-Nam Huynh, Yoshiteru Nakamori, Hiroakira Ono, Jonathan Lawry, Vladik Kreinovich, Hung T. Nguyen (Eds.)
Interval / Probabilistic Uncertainty and Non-classical Logics, 2008
ISBN 978-3-540-77663-5

Ewa Pietka, Jacek Kawa (Eds.)
Information Technologies in Biomedicine, 2008
ISBN 978-3-540-68167-0

Ewa Pietka, Jacek Kawa (Eds.)

Information Technologies in Biomedicine



Springer

Editors

Prof. Ewa Pietka
Silesian University of Technology
Faculty of Automatic Control
Electronics and Computer Science
ul. Akademicka 16
44-100 Gliwice
Poland

Jacek Kawa
Silesian University of Technology
Faculty of Automatic Control
Electronics and Computer Science
ul. Akademicka 16
44-100 Gliwice
Poland

ISBN 978-3-540-68167-0

e-ISBN 978-3-540-68168-7

DOI 10.1007/978-3-540-68168-7

Advances in Soft Computing

ISSN 1615-3871

Library of Congress Control Number: 2008926730

©2008 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 in acid-free paper

5 4 3 2 1 0

springer.com

Preface

Information Technologies do not recognize borderlines between disciplines. This research area rather follows the tradition of interdisciplinary cooperation which requires attention to be given to the needs of other people. In information technologies in biomedicine three very different and specific partners are to work together. Patients needs are recognized by physicians who collaborate with scientists and engineers. This defines the goal of our research which is to satisfy the functional requirements of authorized physicians for the benefit of the patients.

In this book, members of the academic society of technical and medical background present their research results in order to bridge the gap between information technologies and clinical medicine. Researchers who explore the area of medical imaging, biomedical signals, and biotechnology discuss various innovations that change the way of solving problems. New approaches to data processing and recognition increase the efficiency of medical diagnostic and treatment and permit more information to be extracted from the acquired data. The information is then put at the disposal of authorized physicians at the place and at the time it is required, in a format adapted to the specific needs of the physician and the patient.

An extended area is covered by the articles. It includes biomedical signals, medical image processing, recognition, and understating, text processing and natural language analysis, telemedicine, computer-aided diagnosis and treatment, biomaterials. Papers present various theoretical approaches and new methodologies based on fuzzy sets, mathematical statistics, morphological methods, fractals, wavelets, syntactic methods, artificial neural networks, graphs and many others.

We would like to express our gratitude to all paper reviewers as well as the authors who have contributed their original research papers.

Gliwice
June 2008

Ewa Piętka
Jacek Kawa

Contents

Part I: Invited Papers

Exploring the Knowledge of Human Expert beyond His Willing Expression <i>Piotr Augustyniak</i>	3
Application Problems of Implants Used in Interventional Cardiology <i>Zbigniew Paszenda</i>	15
Computer Enhanced Orthopedics <i>Wojciech Glinkowski</i>	28
Computer-Aided Diagnosis: From Image Understanding to Integrated Assistance <i>Artur Przelaskowski</i>	44

Part II: Image Processing and CAD

Biomedical Structures Representation by Morphological Spectra <i>Juliusz L. Kulikowski, Małgorzata Przytułska, Diana Wierzbička</i>	57
Medical Image Analysis Using Potential Active Contours <i>L. Pieta, A. Tomczyk, P.S. Szczepaniak</i>	66
Potential Active Contours – Basic Concepts, Mechanisms and Features <i>A. Tomczyk, L. Pieta, P.S. Szczepaniak</i>	74
Fractal Magnification of Medical Images <i>Jan Kwiatkowski, Wojciech Walczak</i>	85

Fuzzy Clustering in Segmentation of Abdominal Structures Based on CT Studies <i>Wojciech Więclawek, Ewa Piętka</i>	93
The Clusterization Process in an Adaptative Method of Image Segmentation <i>Aleksander Lamza, Zygmunt Wrobel</i>	105
Content-Based Indexing of Medical Images for Digital Radiology Applications <i>Piotr Boninski, Artur Przelaskowski</i>	113
Shape and Texture Feature Extraction for Retrieval Mammogram in Databases <i>Ryszard S. Choraś</i>	121
Mathematical Morphology (MM) Features for Classification of Cancerous Masses in Mammograms <i>Konrad Bojar, Mariusz Nieniewski</i>	129
Stroke Display Extensions: Three Forms of Visualization <i>Artur Przelaskowski, Katarzyna Sklinda, Grzegorz Ostrek</i>	139
Automated Fuzzy-Connectedness-Based Segmentation in Extraction of Multiple Sclerosis Lesions <i>Jacek Kawa, Ewa Piętka</i>	149
Computer-Interactive Methods of Brain Cortical Evaluation <i>Anna Czarnecka, Marek J. Sasiadek, Elzbieta Hudyma, Halina Kwasnicka, Mariusz Paradowski</i>	157
Automatic Registration of MRI Brain <i>Piotr Zarychta, Anna Zarychta-Bargiela</i>	165
Magnetic Resonance Image Classification Using Fractal Analysis <i>Karol Kuczyński, Paweł Mikotajczak</i>	173
Application of MLBP Neural Network for Exercise ECG Test Records Analysis in Coronary Artery Diagnosis <i>Kamil Stefko</i>	179
Volumetric Analysis of Tumours and Their Blood Vessels <i>Rafał Henryk Kartaszyński, Paweł Mikolajczak</i>	184
Pre- and Postprocessing Stages in Fuzzy Connectedness-Based Lung Nodule CAD <i>Paweł Badura, Ewa Piętka</i>	192

Modeling and Simulation of Airway Tissues Stresses during Pulmonary Recruitment <i>Bożena Kuraszkiewicz</i>	200
Compression of Bronchoscopy Video: Coding Usefulness and Efficiency Assessment <i>Artur Przelaskowski, Rafal Jozwiak</i>	208
Fuzzy Rule-Based System for the Diagnosis of Laryngeal Pathology Based on Contact Endoscopy Images <i>Wojciech Tarnawski, Jacek Cichosz</i>	217
Synthesis of Medical Images in the Domain of Melanocytic Skin Lesions <i>Zdzisław S. Hippe, Jerzy W. Grzymala-Busse, L. Piątek</i>	225
Identification of Layers in a Tomographic Image of an Eye Based on the Canny Edge Detection <i>Robert Koprowski, Zygmunt Wrobel</i>	232
<hr/>	
Part III: Signal Processing	
<hr/>	
Diagnostic Quality-Derived Patient-Oriented Optimization of ECG Interpretation <i>Piotr Augustyniak</i>	243
Projective Versus Linear Filtering for Repolarization Duration Measurement <i>Marian Kotas</i>	251
An Application of Robust Kernel-Based Filtering of Biomedical Signals <i>Tomasz Pander</i>	259
Weighted Averaging of ECG Signal Using Criterion Function Minimization <i>Alina Momot</i>	267
Empirical Bayesian Approach to Weighted Averaging of ECG Signal Using Cauchy Distribution <i>Alina Momot, Michał Momot</i>	275
An Approach to Estimation of the Angular Eye-Ball Speed Based on the EOG Signal <i>Tomasz Przybyła, Tomasz Pander, Robert Czabański, Norbert Henzel</i>	283

Ensuring the Real Time Signal Transmission Using GSM/Internet Technology for Remote Fetal Monitoring <i>Krzysztof Horoba, Janusz Wrobel, Dawid Roj, Tomasz Kupka, Adam Matonia, Janusz Jezewski</i>	291
Prediction of Newborn Sex with Neural Networks Approach to Fetal Cardiotocograms Classification <i>Michal Jezewski, Robert Czabanski, Krzysztof Horoba, Janusz Wrobel, Janusz Jezewski</i>	299
Coping with Limitation of Bedside Measurement Instrumentation for Reliable Assessment of Fetal Heart Rate Variability <i>Janusz Wrobel, Janusz Jezewski, Krzysztof Horoba</i>	307
Relationships between Isopotential Areas in EEG Maps before, during and after the Seizure Activity <i>Hanna Goszczyńska, Marek Doros, Leszek Kowalczyk, Krystyna Kolebska, Stanisław Dec, Ewa Zalewska, Jan Miszczak</i>	315
Assessment of Uterine Contractile Activity during a Pregnancy Based on a Nonlinear Analysis of the Uterine Electromyographic Signal <i>D. Radomski, A. Grzanka, S. Graczyk, A. Przelaskowski</i>	325
<hr/>	
Part IV: Biotechnology	
<hr/>	
Use of Computer System for Cell Hybridisation in Biotechnology and Medicine <i>Andrzej Dyszkiewicz, Paweł Poleć, Jakub Zajdel, Damian Chachulski, Bartłomiej Pawlus</i>	335
Clustering as a Method of Image Simplification <i>Anna Korzyńska, Mateusz Zdunczuk</i>	345
Application to Estimate Haplotypes for Multiallelic Present-Absent Loci <i>Robert Nowak</i>	357
Detection of Mitotic Cell Fraction in Neural Stem Cells in Cultures <i>Anna Korzyńska, Marcin Iwanowski</i>	365
Protein Molecular Viewer for Visualizing Structures Stored in the PDBML Format <i>Dariusz Mrozek, Andrzej Mastej, Bożena Małysiak</i>	377

Fuzzy Support Vector Machine for Genes Expression Data Analysis	
<i>Joanna Musiol, Agnieszka Więclawek, Urszula Mazurek</i>	387
Predictive Performance of Top Differentially Expressed Genes in Microarray Gene Expression Studies	
<i>Henryk Maciejewski</i>	395
A Study on Diagnostic Potential of a Computer-Assisted System for Identification of Neoplastic Urothelial Nuclei from the Bladder	
<i>A. Dulewicz, D. Piętka, P. Jaszczak</i>	403
<hr/>	
Part V: Data Analysis	
<hr/>	
Control of Hand Bioprosthesis Via Sequential Recognition of Patient's Intent Using Combination of Fuzzy Sets and Dempster-Shafer Theory	
<i>Marek Kurzynski, Andrzej Wolczowski, Mariusz Topolski</i>	421
Matching Knowledge and Evidence in a Model of Medical Diagnosis	
<i>Ewa Straszeka</i>	429
Nonparametric Regression for Analyzing Correlation between Medical Parameters	
<i>Malgorzata Charytanowicz, Piotr Kulczycki</i>	437
Experiments on Linear Combiners	
<i>Michal Wozniak</i>	445
Processing of Missing Data in a Fuzzy System	
<i>Sylvia Pospiech-Kurkowska</i>	453
Knowledge-Based Decision Hybrid System for the Doctor's Work Support	
<i>Zbigniew Buchalski</i>	461
Features for Text Comparison	
<i>Marek Krótkiewicz, Krystian Wojtkiewicz</i>	468
Possibility of Use a Fuzzy Loss Function in Medical Diagnostics	
<i>Robert Burduk</i>	476

An Application of a Generalized Additive Model for an Identification of a Nonlinear Relation between a Course of Menstrual Cycles and a Risk of Endometrioid Cysts
Dariusz Radomski, Zbigniew Lewandowski, Piotr I. Roszkowski 482

Recognition of the Ventilatory Response to the Intermittent Chemical Stimuli in Awake Animals
Beata Sokolowska, Agnieszka Rekawek, Adam Jozwik 488

Part VI: Multimedia

Telesfor – Telemedical Real-Time Communication Support System
Jerzy Błaszczczyński, Bartłomiej Prędkie, Roman Słowiński 497

Multimedia Program for Teaching Autistic Children
Joanna Marnik, Magdalena Szela 505

Multimedia System for Accessible Distant Education
Dominik Spinczyk, Piotr Brzoza 513

Part VII: Biomechanics

Biomechanical Behaviour of Double Threaded Screw in Tibia Fixation
Witold Walke, Jan Marciniak, Zbigniew Paszenda, Marcin Kaczmarek, Jerzy Cieplak 521

Biomechanical Analysis of Lumbar Spine Stabilization by Means of Transpedicular Stabilizer
Jan Marciniak, Janusz Szewczenko, Witold Walke, Marcin Basiaga, Marta Kiel, Ilona Mańka 529

FEM Analysis of the Expandable Intramedullar Nail
Wojciech Kajzer, Anita Krauze, Marcin Kaczmarek, Jan Marciniak 537

Biomechanical Analysis of Plate for Corrective Osteotomy of Tibia
Jan Marciniak, Marcin Kaczmarek, Witold Walke, Jerzy Cieplak 545

Kinematic Analysis of Complex Therapeutic Movements of the Upper Limb
R. Michnik, J. Jurkojć, Z. Rak, A. Mężyk, Z. Paszenda, W. Rycerski, J. Janota, J. Brandt 551

Influence of Model Discretization Density in FEM Numerical Analysis on the Determined Stress Level in Bone Surrounding Dental Implants	
<i>Jarosław Żmudzki, Witold Walke, Wiesław Chladek</i>	559
Computer Simulations of Electric Properties of Organic and Non-organic Compounds	
<i>P. Janik, M.A. Janik, Z. Wrobel</i>	568
Author Index	573

Exploring the Knowledge of Human Expert beyond His Willing Expression

Piotr Augustyniak

AGH University of Science and Technology,
30 Mickiewicza Ave. 30-059 Krakow Poland
august@agh.edu.pl

Summary. The paper discusses the alternative method of medical experts participation in technical inventions for medicine. Blind tests and various statistic-based correlations of human and automatic interpretation results are commonly used today. Our paper postulates a deeper insight into the expert performance in order to better understanding and simulating his reasoning in the software. The benefit is twofold: the measurement is objective and the closer simulation of human reasoning yields better performance in case of unexpected input. Although the area of application is the very broad intersection of medicine and technology, we focus on the automatic ECG interpretation, and propose the agile software featuring a human-like behavior. Two examples of experiments aimed at extraction of some aspects of ECG interpretation knowledge are also included in the presentation.

1 Introduction

1.1 Rising a Need for New Knowledge

Information technology not only supports the medical practice of today, but also makes new challenges and opportunities stimulating the progress in medicine. The Holter ECG recorders or Computed Tomography may be recalled as first-hand examples of such inventions. In AGH-UST Biocybernetic Lab. we designed and prototyped a wearable ECG recorder-interpreter designed for a wireless cardiology-based surveillance network. Unlike the currently marketed systems, it continuously adapts the ECG signal interpretation process to several prioritized criteria of medical and technical nature. The process is designed as distributed and is performed partially by separated thread on the supervising server (network node) and partially by the agile software of the remote recorder [1]. Important novelty is also the use of digital wireless link in a bi-directional mode for patient and device status reporting but also for management and control of the remote software, requests for adaptation of report contents and data priority and reloading of software libraries as necessary. Such adaptive system yields unprecedented personalization and diagnosis-oriented processing and thus better simulates the seamless presence of a cardiologist. Until today the prototype brings rather scientific challenges, revealing new unexploited areas present in

clinical practice but not covered by the standards, recommendations or guidelines [12].

1.2 Current Validation Rules and Their Limitations

The standards of ECG interpretation quality assessment [6] require the values of the diagnostic results to fall within a specified tolerance range around the value believed to be true. Such “true” reference is usually estimate by averaging the response of independent human or software experts. This approach has two drawbacks:

- it is based on the similarity of results and not of reasoning making even good automatic interpreters useless in case of unexpected input,
- it is optimal for a hypothetical “average” patient a not for a particular person, because of not considering the intra-patient variations.

1.3 Objective Measurement of Behavior

All measurement techniques in technical, economic and social sciences assume the less-possible extent of influence of data acquisition process on the observed phenomena [8]. Technical measurements express this idea as non-energetic information transmission. For this reason, the investigation of the medical expert knowledge using his willing expression may not be accepted as the objective measure. Although assumed not to be biased intentionally, it is influenced by several mental factors. Two principal are: memorization and verbalization.

Memorization uses the short-term memory and a part of human attention to capture his own behavior. The behavior is thus not spontaneous as it were naturally and the auto-observation usually implies subconscious auto-restriction. In result the memorized facts are altered and not complete.

Verbalization is necessary to express the memorized knowledge with a limited set of tokens belonging to a specified vocabulary. Such dictionary depends on the language used, but is also influenced by subjective preferences of the speaker. Therefore the output of an interview with an expert concerning his own reasoning may not be considered more seriously as discrete, incomplete and inaccurate impressions.

Fortunately, the interview as a research methodology has many alternatives, among of them the experiment. From this point of view, however, the originality of our approach is that medical experts are proposed to be subjects in our experiment. The presentation of this innovative idea will be developed throughout this paper.

2 Methodology

2.1 Human Expert as Experiment Subject

Lets take the analogy form the medical diagnostic process. It usually starts with the interview, but commonly needs supplementary tests providing objective measurements of various diagnostic parameters. The patient, assuming he

is a highly cooperative proprietor of the information, is not able to estimate several important facts about himself without specialized sensors and measurement methodology. In the scheme presented above, we postulate to replace the patient by the medical expert at work and to use specialized interdisciplinary technologies aimed at extracting the milestones and foundations of decision-making path. We are conscious, that similarly to the health information, this is a potentially very sensitive area and employ all ethical guidelines to the management of the experimental results. As a principle, we don't make any judgment about the correctness of the results and we keep all demographical data of the experts involved in a separate database.

In order to fulfill the requisite of an objective measurement to the maximum, we should not inform experts about their participation in the experiments. That may be feasible with the video surveillance of people on the street, but not in case of medical task performance. The experts willingly accept the participation in the experiment, as they were informed that they will be observed at work in several manners. Some of them were clearly visible, some others remain undisclosed. The applied measurement techniques should be selected with regard of the efficiency, but also the experimental setup should reproduce as close as possible the natural working environment of the expert.

2.2 Knowledge Exploration Techniques

The most common technique for the exploration of human behavior involves image acquisition and sequences analysis. As the interpretation of ECG signals has no kinetic background, we do not apply the video sequence analysis, however some sessions were video-recorded and the records were found helpful for validation of the session flow, identifying obstacles and interfering events. For the reason of commodity, we also do not apply complicated techniques of brain monitoring. The fMRI [4], although revealing the physiologically evoked parts of the brain is not able to reproduce the reasoning, and does not allow to simulate the usual working environment. The EEG may probably be used in human in course of the visual ECG interpretation process, but, here again long preparation, relatively poor repeatability and no representation of the cognitive process motivated us to focus on more appropriate methods.

The principal method was thus based on the fact, that the ECG interpretation process may be considered as fully visual [2]. By fixing the trace, we may consider the eye gaze trajectory as representative to the reasoning process. In fact, the observation of the trace (or any object) by the human consists of two mental processes running in parallel and interchanging information in restricted time windows [7]. Due to the very limited visual field, after the image is acquired in the human retina, the interpretation starts and as soon as acquisition completes, the research for a new focus point begins in order to provide the interpretation with a complementary image. The sight is a principal sense in human and thus eyetracking techniques are commonly used as objective indicators of visual cues (advertising) or reading and linguistic skills (education).

The second knowledge exploration technique is based on the expert-computer interaction. Assuming the computer is already used to visualize the trace for interpretation and the expert interacts with the system with use of a standard mouse, it doesn't require any additional device. After the interpretation is completed, the expert had to express his preferences in a field of selection marks. Unlike in the standard software, these fields order and the initial selection state were controlled by a random generator what prompts the expert to search for most important items at first and to revert the selection as necessary.

3 Pursuit of the Human Eye in Course of the Visual ECG Interpretation

3.1 Perceptual Model Concept

Perceptual models have been recently recognized as valuable tool enriching the visual interaction of human with sophisticated devices [5, 9]. As a perceptual model of a biosignal record we understand a result of statistical processing of scanpaths, analyzed as polygonal curves in context of displayed visual information. The gaze order and fixation time correspond to the seeking sequence and to the amount of data gathered visually by the observer and thus they represent the diagnostic importance of particular regions in the scene [3, 10]. In the ECG, subsequent events in the cardiac cycle are represented by P, QRS and T waves positions, therefore the wave start- and endpoints were selected as reference time points for the analysis of human foveation sequence aiming at estimating the local density of medical data. Assuming the observer is properly engaged in the trace inspection, the gaze is controlled instinctively and the eyeglobe movements objectively represent the information gathering sequence. The analysis of experts' eyeglobe trajectories captured during the manual interpretation not only reveals regions of particular importance in the signal trace, but also represents the human reasoning involved in the interpretation process. Apart from main interest of our research, the prospective area of applications for eyetrack features captured during the visual inspection of biosignals include:

- objective assessment of cardiologist interpretation skills,
- teaching of the visual interpretation using the guided repetition of expert's scanpath.

3.2 Eye Tracking Method

The infrared reflection-based eyetracker OBER-2 capturing two-dimensional trace of each eye at 750 Hz during the ECG presentation lasting for 8 s was used in visual experiments. The device provides the angular resolution of 0.02 deg and uses time-differential method for the sidelight discrimination. This angle corresponds to a time interval of 30 ms on a standard ECG chart plot (25mm/s) viewed from a typical reading distance (40 cm). The position of both eyes was recorded simultaneously and a custom-developed software detects the dominant

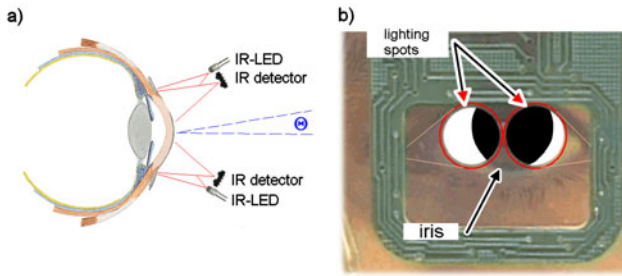


Fig. 1. a) Physical principle and b) technical details of the infrared reflection-based eyetracker OBER-2

eye which trace is used to determine the electrocardiogram conspicuity. Figure 1 displays the physical principle of the differential infrared reflection-based eyetrack acquisition.

3.3 Experiment Setup and Participants

The total of 17 experts (12 ± 4 years of experience) volunteers accepted the invitation to the laboratory for the visual experiment. All observers were asked to complete the statistical questionnaire on their ECG experience and possible eyesight defects before attempting to the visual task. The ECG traces were randomly selected for interpretation from CSE recordings [13] and were presented as bitmaps on a 17 inch CRT monitor. The display simulated a conventional 12-leads paper recording (fig. 2). The reading distance was controlled with use of a chin support set 40 cm apart from the display center. Each ECG trace presentation was interlaced with the fixation point in the center of the display. The reference wave borders, although not displayed, provided the cardio-physiological context for the scanpaths analysis. The horizontal axe of the scanpath is projected on the temporal progress

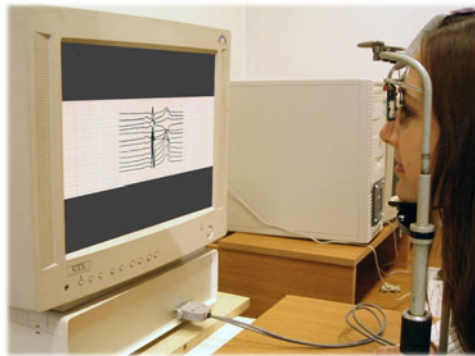


Fig. 2. The expert volunteer performing a visual task on ECG interpretation

of heart cycle, represented by positions of wave borders. Piecewise integration of scanpath time allows to estimate for each cardiac component the amount of information it contributes to the final diagnosis.

3.4 Scanpath Signal Processing

Each visual experiment provides a four-column matrix representing raw eyeglobe coordinates at the evenly spaced time points [11]. Prior to the ECG traces investigation, the calibration rectangle is displayed and the observer is asked to gaze at its corners. The gaze points corresponding to the corners are identified in the eyetrack and help in calculating the display-relative coordinates from the A/D converter output. The further signal processing routines were developed in Matlab with regard to the aims of visual experiments. Main stages of this calculation are performed on the pre-detected dominant eye trace and include:

- the initial idle time t_i and the interpretation task completion time t_e were detected in the scanpath,

$$\forall_{t < t_i} \sqrt{\Delta x_t^2 + \Delta y_t^2} < \epsilon \quad (1)$$

$$\forall_{t > t_e} y_t > m \quad (2)$$

where ϵ is the noise level and m is the maximum vertical screen coordinate

- using a set of reference wave borders S_{\min}^i, S_{\max}^i provided in the CSE database, each foveation point P in the scanpath was qualified as corresponding with the particular ECG sections i ,

$$P \subset S^i : S_{\min}^i \leq p_x < S_{\max}^i \quad (3)$$

- the number and duration D^i of foveation points was integrated separately for each ECG section i in all ECG displays,

$$D^i = \sum_{t=t_i}^{t_e} P(t) \subset S^i \quad (4)$$

- the contribution of each section's conspicuity was referred to the total observation time.

$$C^i = \frac{D^i}{t_e - t_i} \quad (5)$$

The intrinsic variability of waves' length does not influence the result, since the foveation points are referred to ECG fiducial points and not directly to the ECG time. Apart from the waves conspicuity statistics, the processing reveals the perceptual strategy related to main stages of the ECG interpretation process. The principle of strategy description is the identification of:

- most attractive points coordinates,
- their gaze order in context of the ECG time and displayed ECG leads.

These parameters were chosen as most representative to the global density of diagnostic information distribution in the heart cycle and to the information priority required by a diagnostic decision scheme followed intuitively during the manual ECG interpretation by the human expert.

3.5 Results of the Human Eye Pursuit

The statistical parameters of all visual experiment results are summarized in table 1. Figure 3 displays an example of eyeglobe trajectory over a 12-lead ECG plot together with the corresponding bar graph of attention density.

The results in table 1 prove the common belief about irregular distribution of medical data in the electrocardiogram. However, main novelty here is the quantitative assessment of expert's attention density and its variations in the heart cycle. As much as 38 percent of information in the signal is represented in the

Table 1. Results of ECG inspection scanpaths analysis

Parameter	Unit	Observers Experts
idle time	ms	73 ± 55
interpretation time	s	5.5 ± 1.5
P wave foveation	%	23 ± 12
PQ section foveation	%	7 ± 5
QRS wave foveation	%	38 ± 15
T wave foveation	%	18 ± 10
TP section foveation	%	14 ± 5
max. attention density	s/s	21.0
min. attention density	s/s	1.9

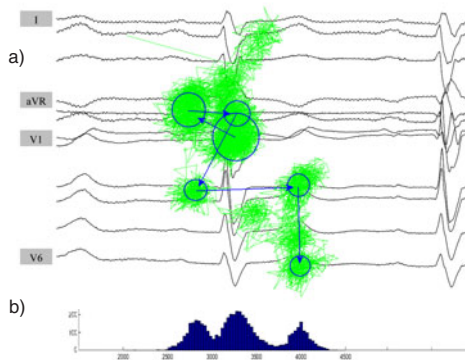


Fig. 3. a) The example of expert's eyeglobe trajectory over a 12-lead ECG plot (CSE-Mo-001); the circle diameter represents foveation time, b) corresponding bar graph of the attention density

QRS complex attracting the experts' gaze to this relatively short (105 ± 23 ms) section. Despite the considerable length of the baseline (278 ± 115 ms), only 14 percent of gaze points fall to this section confirming its little diagnostic significance. Comparing minimum and maximum attention density values we found interesting that this value varies over 10 times, what represents high expert concentration on most informative part of the image. The second group of result were derived by the analysis of perceptual strategy. Figure 3a displays an example of the strategy over a 12-lead ECG plot and table 2 summarizes the corresponding strategy description parameters. For studies on perceptual strategy repeatability we selected electrocardiogram images investigated by at least two observers. By comparing the positions and gaze order of five most important foveation points in the scanpaths we found several different strategies applied by the experts. Having no means to assess them, we only rank them by frequency and observe, that the similarity between two experts may be expected with the probability of 37%. This result prove the proper representation of ECG interpretation process in the visual strategy.

Table 2. Quantitative description of the most frequent perceptual strategy

Parameter	Unit	Observers Experts
relative foveation time for the main focus point	%	31 ± 12
number of foveation points		6.1 ± 1.7
foveation points distance	deg.	5.7 ± 2.4
scanpath length to the last foveation point	deg.	34.7 ± 5.1
scanpath duration to the last foveation point	s	3.6 ± 1.3

The scanpath, however, is very sensitive to the voluntary observer cooperation during visual tasks. Poor cooperation or misunderstanding of visual task rules was the main reason for exclusion of 18% of records from the scanpaths statistics. The scanpath statistics and perceptual strategies revealed many differences between cardiology experts concerning the ECG inspection methods. However, all the statistical parameters indicate a very precise and consistent way of information search by experts. Moreover, high variation of first foveation points focus time and distance suggest the hierarchical information gathering reflecting the parallel decisive process.

4 Pursuit of the Human Choice in Course of Diagnostic Result Selection

4.1 Expert's Choice as Indicator of Medical Data Relevance

Following the requisites of objective measurement, the pursuit of human choice for diagnostic results priority was made with use of a hidden poll. In order to

avoid all-inclusive selections, an artificial restriction of resources was applied. It is based on the expected data stream value attributed to each diagnostic parameter. The total available data volume was set as equal to a half of the data volume of all parameters. In such environment, the doctor has to allocate the space first for the most relevant data, and simultaneously exclude the data he or she considers useless. The aim of this investigation was to record and analyze the expert's behavior in order to extract the knowledge about the relative relevance of ECG diagnostic parameters in most frequent diseases. Such hierarchy yields promising advantages in systems with patient-specific adaptation of the interpretation processing.

4.2 Usual Interface with Hidden Poll Functionality

The manufacturer of ECG interpretation software performs a standard technical validation procedure which is followed by clinical usability verification in selected cardiology expert offices. This last step is very important as it is done by medics, able to demonstrate the software usefulness in live conditions. A trivial modification of the commercial ECG interpretation software (Cardioteka ©, Aspel) was a background for experimental studies concerning doctors' choice about the report contents. The default proposal of a final report contents was replaced by a random pre-selection (fig. 4).

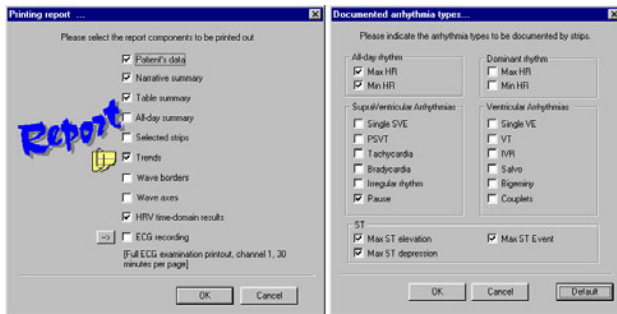


Fig. 4. Example selection screen for the choice on the report contents. Subsequent displays differ by items order and initial selection state.

Once the interpretation is completed, all available report items appear together on the screen and the doctor had to select (deselect) results he or she wish to include in (exclude from) the report contents. The order of selections made and chosen items are memorized with the diagnostic outcome. The survey included 1730 ECG analysis cases and allowed to pursue the cardiologists' preferences in 12 most frequently observed diseases (normal sinus rhythm, sinus tachycardia, sinus bradycardia, probable AV block, ventricular escape beats,

atrial fibrillation, AV conduction defect, myocardial infarction, atrial enlargements, ventricular hypertrophies, left bundle branch block, right bundle branch block). The observations count for these pathologies ranged from 16 to 323 cases. For other 17 pathologies, the occurrence frequency was below 16 in the available population and thus no statistically-justified conclusions may be drawn from.

4.3 Statistical Processing

Finally, the statistical processing of the gathered data was aimed at revealing the knowledge about doctors' preferences. Main steps of the calculations were the following:

- Inclusion or exclusion of a parameter to/from a diagnostic report increases or decreases its relevance accordingly to the expert action. First inclusion is the most relevant, first exclusion is the least relevant etc. Items remaining untouched by the expert are not considered for the hierarchy statistics.
- For each considered disease, the disease-specific hierarchy list was build of diagnostic parameters p , ordered by their frequency F of occurrence at a given position L_C relative to all occurrences at other positions L weighted by their distances $|L - L_C|$:

$$F = \frac{\sum C : L_C = L}{\sum (C \cdot |L - L_C|) : L_C \neq L} \quad (6)$$

- The diagnostic relevance is represented by the weighting coefficient W_p including the rank L and the frequency F :

$$W_p = \frac{F}{L} \quad (7)$$

- Finally, the weighting coefficients were normalized so as they sum to the unity

$$\sum_p W_p = 1 \quad (8)$$

This operation yields a disease-specific vector of weighting coefficients representing the medical relevance of ECG diagnostic parameters (tab. 3)

4.4 Results of the Pursuit of Human-Made Relevance Indication

These studies confirm the common, but poorly justified belief, that for the human expert some diagnostic results are more important than others. The relevance of particular medical parameters highly depends on the known status of the patient. Moreover, assuming that several common diseases may be reliably diagnosed in a fully automated process, our studies result in attributing each disease with a hierarchical list of most adequate diagnostic parameters. That list is very useful as a background of disease-dependent report modification in a distributed cardiac surveillance system. It may also be applicable to a medically-justified estimation of cardiac data quality.

Table 3. Hierarchy matrix of diagnostic parameters in patient state (excerpt); The normalized weighting parameters W_p are displayed

diagnostic parameter	patient status space			
	normal sinus rhythm (111)	persistent supraventricular tachycardia (163)	ST suppression heart muscle ishaemiae (508)	...
heart rate (HR) [1/min]	0.15	0.35	0.22	...
dominant trigger rate [%]	0.25	0.21	0.07	...
PQ section [ms]	0.12	0.17	0.03	...
...

5 Conclusions

The paper presents the idea of behavioral observation and measurements in human, applied to the extraction of the cardiology expert knowledge being a background of the visual signal interpretation and management of parameters. The methodology of undisclosed observation is not new, it is commonly accepted in sociology and medical sciences. It fulfills very well the requisite of objective or unbiased measurement.

Although very advantageous, the method involves ethical issues, and probably such methods should be used under a supervision of ethical commissions similar to other experiments *in vivo*. The human under test being a proprietor of the knowledge and of the performance, has only a limited influence on the information he or she provides to the analytic system. On the other hand, medicine is at the leading edge in the usage of similar experiments in animals and in human and therefore high level of understanding from doctors participants should hopefully be expected.

Acknowledgement. Scientific work supported by the AGH-University of Science and Technology grant No 10.10.120.783.

References

1. Augustyniak, P.: The use of selected diagnostic parameters as a feedback modifying the ECG interpretation. In: Proc. Computers in Cardiology, vol. 33, pp. 825–828 (2006)
2. Augustyniak, P., Tadeusiewicz, R.: Assessment of electrocardiogram visual interpretation strategy based on scanpath analysis. *Physiol. Meas* 27, 597–608 (2006)
3. Boccignone, G.: An Information-theoretic Approach to Active Vision. In: Proc. 11th International Conference on Image Analysis and Processing (2001)

4. Buxton, R.B.: *An Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques*. Cambridge University Press, Cambridge (2002)
5. Dick, A.O.: *Instrument scanning and controlling: using eye movement data to understand pilot behavior and strategies*. NASA CR 3306 (1980)
6. IEC, 60601-2-47. *Medical electrical equipment: Particular requirements for the safety, including essential performance, of ambulatory electrocardiographic systems* (2001)
7. Kasprowski, P., Ober, J.: *Eye Movements in Biometrics*. In: *ECCV Workshop BioAW*, pp. 248–258 (2004)
8. Kimothi, S.K.: *The Uncertainty of Measurements: Physical and Chemical Metrology: Impact and Analysis* (online) (2002)
9. Ober, J.K., Ober, J.J., Malawski, M., Skibniewski, W., Przedpelska-Ober, E., Hryniewiecki, J.: *Monitoring Pilot's Eye Movements during the Combat Flight-The White Box*. *Biocybernetics and Biomedical Engineering* 22(2-3), 241–264 (2002)
10. Pelz, J.B., Canosa, R.: *Oculomotor behavior and perceptual strategies in complex tasks*. *Vision Research* 41, 3587–3596 (2001)
11. Salvucci, D.D., Anderson, J.R.: *Automated Eye-Movement Protocol Analysis*. *Human-Computer Interaction*, vol. 28 (2000)
12. *The standard action in Holter examination*. Polish Society of Noninvasive Electrocardiography (1997)
13. Willems, J.L.: *Common Standards for Quantitative Electrocardiography 10-th CSE Progress Report*, Leuven: ACCO publ. (1990)

Application Problems of Implants Used in Interventional Cardiology

Zbigniew Paszenda

Silesian University of Technology, Institute of Engineering Materials and Biomaterials, ul. Konarskiego 18a, 44-100 Gliwice, Poland
zbigniew.paszenda@polsl.pl

Summary. The paper discusses application issues of using the metal implants for treatment of the cardiovascular diseases. The analysis of the biophysical conditions of the heart-coronary vessels system has been used to distinguish the tissue environment properties which should be compatible with properties of the metal biomaterial and stent surface. The need to determine the correct quality and service properties of the coronary stents has been indicated, which refer first of all to their design form, physical and chemical properties of the metal biomaterial and its surface. Based on that the Author of the work has proposed his own methodology for forming and controlling the service properties of the stents. It takes into account the required relationships between structure, and mechanical properties of the stent biomaterial, and the physical and chemical properties of its surface - adjusted to the specific features of the cardiovascular system.

1 Introduction

Employment of the intravascular implants called *stents* has become one of the most important achievements of the nineties of the last century in the area of surgical cardiology in the ischaemic heart disease treatment. These implants feature a sort of a metal, small sized, springy scaffolding with the spatial cylindrical design grafted into the critically stenosed coronary vessel to support and simultaneously dilate its active section. The Percutaneous Transluminal Coronary Angioplasty (PTCA) was one of the main coronary heart disease treatment methods used to dilate the vascular lumen in the period before the stents were introduced, apart from the pharmacological therapy and Coronary Artery Bypass Graft (CABG). The idea of such operation was presented first by Dotter and Judkins in 1964. This method was effective mainly for peripheral arteries. Potsmann modified this operation using the catheter ending with delivery balloon for the first time. To dilate the coronary arteries this operation was improved and introduced for the first time by Gruentzig in 1977 at the University in Zurich [1, 2, 3].

In spite of many advantages resulting from introducing the PTCA operation in the ischaemic heart disease treatment it has also some limitations. They include the risk of restenosis (about 30÷50% patients) and of the rapid occlusion of

the coronary artery (about 7% patients) [1, 4, 5, 6]. Introduction of the coronary stents into the clinical practice has been the effect of many years of attempts to overcome these problems. A significant interest in this treatment method has followed the simultaneous publication of two, classic as of today, scientific contributions resulting from the Belgian-Dutch cooperation: BENESTENT (Belgium Netherlands Stent) and STRESS (Stent Restenosis Study) [7]. The authors of these works proved that grafting a stent into the proper location of the coronary system reduces significantly the angiographic restenosis frequency (by about 50%) with patients at risk of the new lesions.

It follows from studies of most works pertaining to use of the coronary stents that their effectiveness is decided mostly by the physical and chemical properties of the implants surfaces. Therefore, the current research is focused mostly on development of a method for deposition of coatings on the metal stents surfaces, which reduce significantly the blood clotting process and ensure their good bio-tolerance in the cardiovascular system tissues environment. Numerous publications in the world literature confirm these activities. However, they present most often the partial research results only (mostly biological ones in the *in vitro* and *in vivo* conditions) which do not make full assessment possible of the fabricated coatings usefulness, e.g., their corrosion resistance or adhesion to the stent surface. Moreover, differentiation of methodologies of the research carried out does not always make it possible to compare results obtained by different authors. The issue of the relevant metal biomaterial selection is also left out in the presented works (chemical composition, microstructure, and mechanical properties) deciding the service properties of the investigated stent. Therefore, fabricating the atombogeneous coatings on stents surfaces should be preceded by selection of the biomaterial with a structure and physical properties taking into account specific features of stents (their miniaturisation, implantation technique) conditioned by the cardiovascular system (mainly by the biochemical, bioelectronic and biomagnetic factors). The quality criteria for the metal biomaterials presented so far do not specify recommendations for this form of implants – stents.

2 Biophysical Conditions of the Heart-Coronary Vessels System

Biophysical conditions of the cardiovascular system result from the possibility of generating the action potentials by the cardiac muscle cells and from the specific features of the coronary vessels system. Mechanisms of generating the action potentials are based on ion- and electric charges transport through its cell membranes. The effect is the flow of the ion electric current (action current) with the varying intensity. Therefore, these currents are responsible for generating the alternating electrical field in the living organism. So, the beating heart may be considered to be an electric dipole changing in time. This macroscopic dipole is a resultant of many microscopic dipoles, as the activated cardiac muscle fibres are assumed to be. The activated part of the muscle fibre is the negative-, and

the inactive part – the positive pole of such dipole. The resultant of these dipoles at a given moment features the main electrical vector of the heart [8].

Electrical excitations in the cardiac muscle cells are also the main source of the organism's magnetic field. The magnetic field, in the simplified mathematical description, is considered to be generated by the current dipole or a set of dipoles placed in the isotropic conducting material with the constant conductance. The density values of the ion currents passing during the heart work are small. Therefore, the real values of the generated magnetic field induction measured outside of the organism are also small and are several picoteslas only [1].

Changes of the electrical and magnetic fields are orthogonal to each other. They are considered separately at low frequencies of the emitted electromagnetic waves. Emission of the electromagnetic waves with low frequency from the ELF (Extremely Low Frequency) range takes place during heart action. Therefore, to evaluate its action, separate analysis methods for its electrical- (electrocardiography – ECG) and magnetic (magnetocardiography – MCG) fields changes were proposed.

The need results from the analysis carried out to adjust the physical properties (electrical and magnetic) of the metal implants to the specific features of the cardiovascular system. Interference with such system by grafting a metal implant should not affect processes connected with generation and propagation of the action potentials in the tissues. Moreover, appearance of an implant with the ferromagnetic properties would not leave the electromagnetical processes unaffected. The effect of such implant might turn out to be even more harmful if the effect of the external electromagnetic field, to which its user may be subjected, is taken into account.

The biophysical properties of the coronary vessels affect strongly the coronary blood flow process. The fundamental role in ensuring the relevant properties of the vessels is played by their internal layer – endothelium. That is just the endothelium cells that produce many substances (mostly NO) affecting, among others, the active tension state of the vascular muscles and their atrombogeneous properties of their internal walls. Therefore, the development of the disease process and in consequence the faulty oxygen delivery to the cardiac muscle cells are eventually dependant on the proper flow of processes of synthesis and releasing the biologically important elements, as well as on phenomena on the endothelium surface – flowing blood interface. Therefore, the grafted intravascular implant should be characteristic of such physical and chemical properties of its surface that it would not initiate development of the disadvantageous reactions disturbing additionally functioning of the endothelium (apart from the originated already disease process).

Haemostasia induced by presence of the metal implant is one of the negative phenomena occurring in the cardiovascular system environment [1]. The process of blood interaction with the implant materials is still not fully understood. It is generally assumed that due to blood contact with the *artificial* implant surface adsorption of proteins (mostly of fibrinogen) occurs first. In case when the adsorbed fibrinogen undergoes the denaturisation process, the next platelet

and plasma blood clotting factors get activated in a cascading way. This in consequence leads to development of a clot.

One of the mechanisms explaining the nature of the clotting process initiation is based on the energy band diagram [9, 10, 11]. It was found out based on investigations of Gutmann and his associates that fibrinogen has the electron structure characteristic of the semiconductor materials. The width of its forbidden band is 1.8 eV. Its valence- and conduction bands are at 0.9 eV below or above Fermi level respectively. Therefore, the protein transformation process from its inactive form (fibrinogen) into the active one (fibrin) may be connected with the electrochemical reaction occurring between the protein and the material surface being in contact with blood. Electrons from the fibrinogen valence band transferred, e.g., to implant material cause disintegration of protein. The consequence is transformation of the protein into a monomer and fibrin peptide. Next the process of their networking occurs leading to the irreversible form of a thrombus. Therefore, it seems purposeful to carry out modification of the physical properties of implant materials by their surface treatment. Fabrication of a layer on implant surface with the high corrosion resistance and semiconductor or dielectric properties may effectively impede transferring electrons from the fibrinogen valence band. This may, in consequence, feature the effective method to limit the blood clotting process due to contact with the grafted implant's surface.

3 Conditions of Using Metal Biomaterials for Coronary Stents

Using the coronary stents in the ischaemic heart disease treatment is possible due to experience gathered with using the metal materials for implants in the orthopaedic- and maxillofacial surgery, alloplastics of joints, and in cardiosurgery. Analysis of stents used in clinical practice makes it possible to specify the following material groups used for their fabrication: Cr-Ni-Mo austenitic steel, alloys of Ni-Ti and platinum, Co-Cr-Ni-W, and tantalum [1, 12, 13, 14]. The Cr-Ni-Mo austenitic steel grades are used most often for the coronary stents – Table 1 [15]. This group of biomaterials is known and commonly used since many years, mostly for the short-term implants in the injury-orthopaedic-, maxillofacial-, and thoracosurgery.

Coronary stents, albeit made from the Cr-Ni-Mo steel belong to the long-term implants. Therefore, lastly the interest has grown in this metal materials group because of its applications for implants contacting blood. This interest is focused mostly on development of various coatings technologies with the atrombogeneous properties, i.e., counteracting the blood clotting process on their surface. However, a few works only are dedicated to the problem of forming the structure and physical properties of the Cr-Ni-Mo steel. Quality criteria pertaining their use for various implants are included in the relevant legislative acts [1, 15, 16, 17]. However, the presented recommendations do not take into account the specific problems connected with using this steel for the coronary stents and do not refer

Table 1. Chemical composition of the Cr-Ni-Mo steel used on implants [15]

Standard	Steel grade	Concentration of elements, %										
		C	Si	Mn	P	S	N ₂	Cr	Mo	Ni	Cu	Fe
ISO5832-1	D	max 0.030	max 1.0	max 2.0	max 0.025	max 0.01	max 0.10	17.0÷ 19.0	2.25÷ 3.50	13.0÷ 15.0	max 0.50	rest
	E	max 0.030	max 1.0	max 2.0	max 0.025	max 0.01	0.10÷ 0.20	17.0÷ 19.0	2.25÷ 3.50	14.0÷ 16.0	max 0.50	rest
ASTM F-139-96	AISI 316L	max 0.030	max 0.75	max 2.0	max 0.025	max 0.01	max 0.10	17.0÷ 19.0	2.0÷ 3.0	12.0÷ 14.0	max 0.50	rest

to their geometrical features and discussed above conditions of the cardiovascular system. Therefore, there is a need to specify clearly the qualitative criteria for this group of the implant materials.

The qualification base for the metal biomaterials is, first of all, determining their chemical composition and structure. Based on the many years long investigation of their biotolerance in the environment of tissues and body fluids, the ranges were determined for the particular chemical elements ensuring the paramagnetic austenitic structure and good pitting corrosion resistance of the steel – Table 1. High requirements connected with service properties of the Cr-Ni-Mo steel grades intended for implants force using smelting methods ensuring their relevant metallurgical purity. However, the non-metallic inclusions remaining after these processes have a significant effect on the service properties of products. This effect depends on the shape, geometrical features, and homogeneity of their distribution. Moreover, during the plastic treatment some of these inclusions are subject to deformation, which is the cause for the mechanical properties anisotropy.

This issue assumes the particular importance when referred to the coronary stents because of the miniature sizes of this type of implants. This forces the need to use steel with the good metallurgical quality characteristic of the minimum amount of the non-metallic inclusions with big dispersion and fine austenite grains. Such structure ensures also a good corrosion resistance, especially in the environment of tissues and body fluids. The methods (mostly comparative ones) recommended in the standard [15] and criteria for the structure quality assessment are useless for determining the quality of steel for coronary stents. Therefore, it seems necessary to specify the exact quantitative relations using the automatic image analysis methods for this type of medical products connected with high risk for the user. Applications of such techniques make measurement possible of the commonly used stereological parameters of a single particle, e.g., its volume, transverse section area, maximum and minimum chord length. Moreover, features like number of particles in a unit volume (area) and range of their sizes are connected with the notion of distribution of dimensions of the non-metallic inclusions. The effect of such analysis may be, therefore, elimination

of the template scale used so far according to standards, or assigning them the particular geometrical parameters of the non-metallic inclusions.

The austenite grain size is an important issue pertaining to the metal biomaterials for stents. It is the main structural parameter affecting strongly the mechanical properties of metal materials. Hall-Petch equation is the most known relationship determining the effect of the average grain diameter on the lower yield point σ_y [1]. The grain size affects significantly also the fatigue strength σ_f of the constructional materials. Results of many investigations confirm that the fatigue strength grows as the grain size gets smaller. Therefore, the analysis above - taking into account the service conditions of the coronary stents (cyclic loading) and their miniaturisation - indicates to the need to use biomaterials with the fine-grained structure. According to requirements posed by the standard it is assumed for the Cr-Ni-Mo steel that the grain size should not exceed the one corresponding to template $G=4$ [15]. Assuming this value as the grain size criterion the average grain size is $d_m=0.088$ mm. Analysis of the geometrical features of stents manufactured nowadays shows that stent thickness is in the $g=0.06\div 0.14$ mm range. This means that a single grain is contained on the transverse section of the stent wall with the thickness of $g=0.09$ mm. In this situation the undoubtedly low ductility of the implant material renders it useless as early as at its implanting stage. This may also be the cause of the unsatisfactory stent life, which is of the utmost importance when the long-term implants are concerned. Therefore, one can state that dispersion of the metal biomaterials structure features the key issue for this form of implants and is not fully determined by the standard recommendations used nowadays.

Selection of the mechanical properties of the metal biomaterial is an important problem in the process of forming the service properties of implants. This problem has been discussed rather widely in literature referring to implants used in the orthopaedic- and maxillofacial surgery, as well as in alloplastics of joints [18, 19, 20]. The optimisation process of the mechanical properties for implants used in the interventional cardiology should be carried out taking into account loads resulting from the implanting technique, which do not occur in their service. This is connected with the necessity to deform the stent permanently to the required diameter to place it in the blood vessel whose patency is being restored. Results of the model testing carried out using the finite elements method are presented in the literature, as there is no possibility to determine the mutual interactions of stents and blood vessels in the investigations in vivo. Having the 3-D model of the stent implanted into the blood vessel and its mechanical properties one can evaluate interactions between these objects. The numerical calculations carried out refer most often to the stress and strain distributions of the particular elements of the assumed system and the blood flow - Fig. 1 [21]–[25]. This makes optimisation possible of the implant's geometrical features and of its biomechanical properties. The degree of strain hardening of the proposed metal biomaterial should be selected so that the determined values of the reduced stress in the stent elements after it expands to the required diameter would exceed its yield point $R_{p0.2}$. The numerical simulations carried out make

it also possible to determine many parameters essential for evaluation of the clinical usefulness of the particular stent forms, e.g., expanded metal surface area and shortening of the stent after its expansion.

Introducing the metal implant into the tissue environment of the cardiovascular system generates additional physical and chemical relationships. They are demonstrated not only by the galvanic effects (initiating the electrochemical processes), but first of all by the electromagnetic ones. This is connected with the action potential generation mechanisms of the cardiac muscle cells and with the action currents flow with varying intensity. One has to take into account in addition the magnetotropism phenomenon (motion reaction induced by presence of the magnetic field) of some tissue structures, and especially of the blood components.

Evaluation of the electrical and magnetic properties of the Cr-Ni-Mo steel is not exposed in the literature. Magnetic properties, in the standard recommendations, were defined by the fact that the Cr-Ni-Mo steel belongs to paramagnets and the ferromagnetic ferrite δ has been eliminated from its structure [15]. The electrical properties of implants may be formed by fabricating layers on their surface which will minimise processes connected with generation and propagation of action potentials in the surrounding tissues.

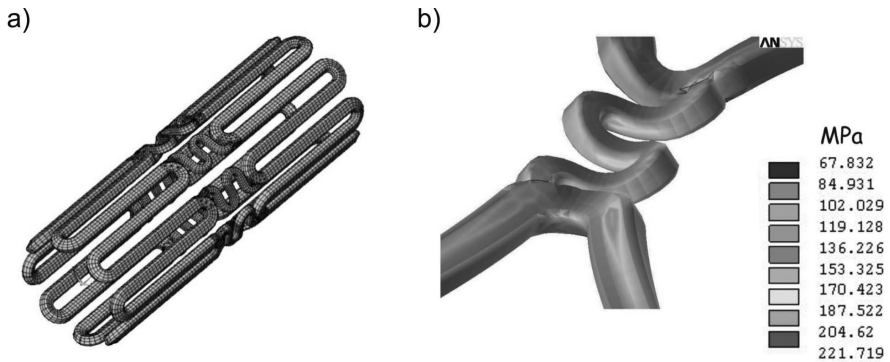


Fig. 1. Results of the numerical analysis of the coronary stent [25]: a) numerical model of the stent, b) stress distribution in the elements of the expanded coronary stent ($p=3$ atm)

4 Forming the Physical and Chemical Properties of the Coronary Stents Surface Layer

Biomaterial introduced into the cardiovascular system may not cause any irreversible damages to the structure of proteins, block activity of enzymes, change composition of the electrolyte, as well as may not damage or release a big number of the morphotic blood components. Moreover, it may not initiate any toxic-, immunological-, or mutagenic reactions. Implanting the metal stent into the vascular system initiates the complex reaction between the blood components and

its surface. Endotelialisation of the implanted stents (covering the stent with endothelium) is a slow process lasting 2÷3 months. This period decides the danger of blood clotting on the implant's surface. After implanting the stent development of a monolayer on its surface begins immediately by the adsorbed blood plasma proteins. As such proteins like fibrinogen or fibronectin have the higher reactivity with the metal implant surface compared to albumin (the most plentiful plasma protein), they may initiate the cascading blood clotting process and adherence of platelets to the implant's surface [7]. Therefore, one can state that the physical and chemical properties of stents surfaces affect intensity of this process. Moreover, they affect the implant – vascular tissue periimplant reactions, deciding the eventual muscle-fibrous proliferation within the tunica intima of the vessel at the stent grafting location.

Roughness of the stent surface features one of the main factors affecting the blood clotting process and reaction of the periimplant tissues. Sheth et al described effects of polishing implants from the Ni-Ti alloy and of Palmaz-Schatz stents from the Cr-Ni-Mo steel in the in vivo tests (in the cardiovascular system of a pig). They have revealed the significant reduction of predisposition to clotting of the polished implants surfaces compared to those that were not polished. Similar tests on rats and pigs were carried out by de Scheerder. Implanting the polished stents from the Cr-Ni-Mo steel he obtained a significant reduction of occurrences of the early thrombosis and limitation of the restenosis process [26, 27]. Results of these investigations indicate to the significant effect of the implants surface topography on the early and late reactions of the stented vessels. Therefore, one can state that ensuring the smooth surface of stents made from the metal biomaterials is the main stage of forming their service properties.

A high rate of development of a layer from the adsorbed plasma blood proteins (mostly fibrinogen) on the stent surface impelled many researchers to search for the more effective methods for limiting this process. A significant interest is observed in using polymers for implants used in the interventional cardiology. Using polymers for implants in the cardiovascular system is not a novel idea. Their good biotolerance in blood environment is mainly stressed, and especially their atombogeneity connected, among others, with the wettability, value and type of the surface electrostatic charge, and the surface conductivity coefficient value of these biomaterials. Polymers are used as the constructional materials for stents fabrication (polymer stents), for composite elements of implants (metal-polymer), and for coatings developed on surfaces of stents.

The number of polymer grades used for protective coatings on surfaces of the metal coronary stents is big and keeps on growing. A big group feature the synthetic non-biodegradable polymers. Their role is development on the stent surface the inactive barrier between the metal stent and the cardiovascular system tissues. Mostly polyurethane, polysiloxane (silicone), and ethylene terephthalate polyester are used for these coatings [28, 29]. Usability of these coatings was evaluated mostly based on the in vivo tests (in the coronary vessels of pigs). Investigations of van der Giessen, de Scheerder, and Fontaine proved their effectiveness in limiting the early blood clotting process on the implants surfaces.

However, these solutions were not effective when limiting the proliferation process of the vessel tunica intima was concerned, as compared to stents without this coating.

Effects of these works shifted the interest to another polymer group. These are the biodegradable polymers (among others lactic polyacids, polyglycolide) [13, 30]. The most advantageous results were obtained when the stent coating material was dilactide (L-lactide). Investigations carried out by Lincoff in the coronary vessels of pigs revealed effectiveness of such coatings both in limiting the clotting process and restenosis.

The idea of using the natural polymer as the coating material for stents which does not initiate the inflammatory process was introduced, among others, by Holmes and Baker [13]. They evaluated usefulness of tantalum stents and those made from the Cr-Ni-Mo steel (of Palmaz-Schatz type) coated with the polyurethane coating into whose surface layer fibrin was introduced. Results of their tests carried out in the *in vitro* and *in vivo* conditions indicate that the coating limits the blood platelets adhesion process and prevents clotting. Moreover, the faster stents endothelialisation was observed. However, employment of such composite coatings is connected with some limitations. Implants should be promptly grafted after introducing fibrin into their surface layer. It is because fibrin does not have a spare adhesion to the substrate.

The next method of forming the physical and chemical properties of stents is heparinisation of their surface [1, 13, 31, 32] It gives the possibility to eliminate administering the antithrombotic drugs to patients in the postoperative period. Heparin contains several anion groups (e.g., carboxyl, sulfane, sulfamide), applying the negative electrostatic charge to the vessels internal surface. This phenomenon was exploited first by Bonan. He used the zig-zag type stents with the deposited heparin coating in his investigations. The results of tests carried out in the *in vivo* conditions did not reveal a positive effect of these coatings on limiting the clotting process. Similar results were obtained by Zidar using the tantalum stents coated with heparin layer. However, Hardhammar, Serruys, Chronos, de Scheerder, and others, in their tests carried mostly in the *in vivo* conditions have observed efficiency of such coating in limiting activation of blood platelets and origination of clots. However, further histomorphometric examinations revealed lack of efficiency of implants prepared in this way in limiting the restenosis process.

Investigations are also ongoing on using the amorphous silicon carbide as coating for the coronary stents [1, 13]. They are connected with the strive to limit the fibrinogen into fibrin conversion occurring on the implant's surface. This type of coating material for stents with the dielectric properties reduces effectively reactivity of its surface in blood environment. The investigations carried out indicate to the good corrosion resistance of such layer and also to the limited capability to initiate activation of blood platelets and aggregation of leucocytes.

The significant progress in treatment of the early and late thrombosis and secondary stenosis of the coronary vessels was achieved by using the drugs releasing stents. The atrombogeneous and anti-inflammatory substances (drugs)

are introduced into the structure of the polymer coatings on stents, which are released into the blood vessel after implanting the stent. Results of the clinical trials presented on scientific sessions of the American Heart Association in Anaheim in 2001 indicate that it is one of the most outstanding achievements in the interventional cardiology since introduction of implants for restoring patency of vessels. Clinical trials were conducted on 44 patients included in the RAVEL programme (the first clinical trial of using the sirolim covered stents for humans) [1, 13, 33, 34]. No restenosis was revealed with angiography for any patient after 12 months from implanting the stent. Also in other works their authors presented investigation results confirming efficiency of such therapy. However, due to the relatively short application period of such stents in clinical practice it is hard to determine clearly efficiency of their many years long service. Therefore, further investigations are needed taking into account not only the type of the drug used but also the type of the drug releasing platform. The drug may be applied directly onto the stent surface or onto the so-called matrix controlling its release. Research is also carried out of the possibility of employing the multilayer coatings using two kinds of drugs A and B separated with the interlayer – Fig. 2.

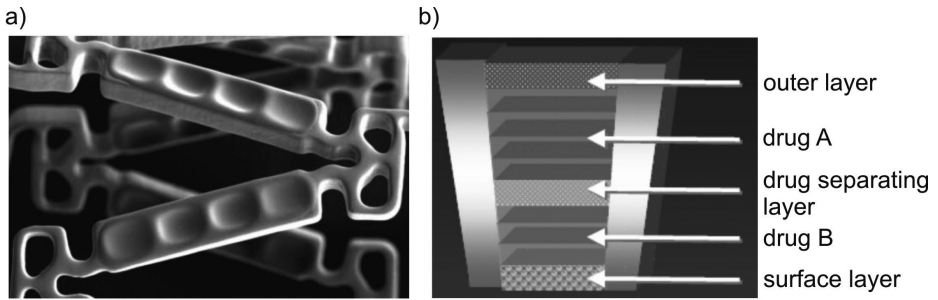


Fig. 2. Prototype of the stent eluting drugs in controlled and multilayer way [13]: a) stent with the so called „baskets”, b) “basket” with the separate layers

Reports were also published about the advantageous effect of the carbon layers on the biotolerance of implants from the metal biomaterials [1, 13, 35, 36]. Within the framework of his research the Author of this work has developed the technology of fabricating the passive-carbon layer with the amorphous structure and verified its final quality in conditions reflecting specifics of using this type of implants. He has demonstrated, based on measurements carried out, that the passive-carbon layer with the nanocrystalline structure fabricated on the coronary stents surface, with the high surface smoothness and dielectric properties ensures fully their pitting corrosion in the implanting and service conditions. The proposed layer does not initiate the stents decohesion processes. Tests of interaction with blood carried out in the *in vitro* conditions and the *in vivo* tests carried out on the experimental animals have confirmed usefulness of the passive-carbon layer fabricated on the coronary stents surface.

5 Summary

Coronary stents have changed fundamentally during the last dozen years or so the methods and effectiveness of the ischaemic heart disease treatment. However, the clinical practice, indicates certain limitations connected mostly with introducing the metal material into the human blood stream. Analysis of the literature data carried out makes it possible to state that these issues are mainly connected with blood clotting on stents surfaces and with restenosis. Much less attention is paid to the problem of selection of the form and geometrical features of stents, as well as to forming the structure and mechanical properties of the metal biomaterial. After all these issues are of the deciding significance during the stent implanting operation and its capability to carry the loads. Also miniaturisation of this form of implants enforces the need to use the biomaterial with the fine-grained structure with the strictly limited content of the dispersive non-metallic inclusions.

Based on the analysis of using the metal biomaterials for the coronary stents and the biophysical conditions of the heart-coronary vessels system the Author of this work has proposed his own methodology for forming and controlling their service properties [1, 2, 35, 36]. The methodology worked out is based on the statement that the required service properties of the coronary stents may be achieved only by forming properly their structure, mechanical properties of the implants biomaterial, and of the physical and chemical properties of their surface, taking into account the specific features of the low-invasive implanting technique and the biophysical conditions of the heart-coronary vessels system.

The methodology algorithm consists of six main stages encompassing – according to the author – investigation steps that have to be carried out [1, 2, 35, 36]:

- analysis of the physiological and biophysical processes of the cardiovascular system – determining the biochemical and biophysical features of the tissue environment that should be taken into account both during forming the microstructure and physical and chemical properties of stents, as well as during specifying their final quality assessment conditions,
- numerical analysis of the stent – coronary vessel system – determining the values of the design parameters characterising the analysed stent form,
- investigation of the metal biomaterial – analysis of its chemical- and phase compositions (determining the geometrical features of the biomaterial's structural components), and its mechanical properties,
- fabrication and surface treatment of the coronary stent,
- investigation of the stent's surface layer structure with methods taking into account the type of the coating used,
- examination of the physical and chemical properties of implants in the in vitro and in vivo conditions.

The developed methodology of forming and evaluation of the final quality of stents is of the interdisciplinary and universal nature. It can be used for the arbitrary forms of implants used in the interventional cardiology.

References

1. Paszenda, Z.: Kształtowanie własności fizykochemicznych stentów wieńcowych ze stali Cr-Ni-Mo do zastosowań w kardiologii zabiegowej. Wydawnictwo Politechniki Śląskiej Gliwice (in Polish) (2005)
2. Paszenda, Z., Duda, B., Wilczek, P.: Investigation of haemocompatibility of the passive-carbon coatings used for improvement of the coronary stents' surfaces. *Engineering of Biomaterials* 26, 3–11 (2003)
3. Paszenda, Z.: Issues of metal materials used for implants in interventional cardiology. *Engineering of Biomaterials* 21, 3–9 (2002)
4. Peng, T., Gibula, P., Yao, K., Goosen, M.: Role of polymers in improving the results of stenting in coronary arteries. *Biomaterials* 17, 685–694 (1996)
5. Lahann, J., Klee, D.: Improvement of hemocompatibility of metallic stents by polymer coating. *Journal of Materials Science: Materials in Medicine* 10, 443–448 (1999)
6. Verweire, I., Schacht, E., Qiang, B., Wang, K.: Evaluation of fluorinated polymers as coronary stent coating. *Journal of Materials Science: Materials in Medicine* 11, 207–212 (2000)
7. Serruys, P., Kutryk, M.: *Handbook of coronary stents*. Martin Dunitz Ltd (1998)
8. Jaroszyk, F.: *Biofizyka*. Wydawnictwo Lekarskie PZWL, Warszawa (in Polish) (2001)
9. Huang, N., Yang, P., Cheng, X., Leng, Y.: Blood compatibility of amorphous titanium oxide films synthesized by ion beam enhanced deposition. *Biomaterials* 19, 771–776 (1998)
10. Chen, J., Leng, Y., Tian, X., Wang, L., Huang, N., Chu, P., Yang, P.: Antithrombotic investigation of surface energy and optical bandgap and haemocompatibility mechanism of titanium oxide thin films. *Biomaterials* 23, 2545–2552 (2002)
11. Huang, N., Yang, P., Leng, Y., Chen, J., Sun, H., Wang, J.: Haemocompatibility of titanium oxide films. *Biomaterials* 24, 2177–2187 (2003)
12. Colombo, A., Stankovic, G., Moses, J.: Selection of coronary stent. *Journal of the American College of Cardiology* 6, 1021–1033 (2002)
13. Marciniak, J., Paszenda, Z., Walke, W., Tyrlik-Held, J.: *Stenty w chirurgii małoinwazyjnej*. Wydawnictwo Politechniki Śląskiej Gliwice (in Polish) (2006)
14. Serruys, P., Rensing, B.: *Handbook of coronary stents*. Martin Dunitz Ltd (2002)
15. ISO 5832-1, *Implants for surgery – Metallic materials – Wrought stainless steel* (2007)
16. EN 12006-3 *Non-active surgical implants – Part 3: Intravascular devices* (1998)
17. EN ISO 14630 *Non-active surgical implants – General requirements* (1997)
18. Popko, J., Szeparowicz, P., Sajewicz, E., Sidun, J.: Biomechanical evaluation of two cervical spine stabilization systems. *Acta of Bioengineering and Biomechanics* 4, 72–79 (2002)
19. Pezowicz, C.: Experimental investigation of cervical spine fixators. *Acta of Bioengineering and Biomechanics* 3, 3–13 (2001)
20. Pozowski, A., Będziński, R., Ścigała, K.: Stress distribution in knee after operative correction of its mechanical axis. *Acta of Bioengineering and Biomechanics* 3, 31–40 (2001)
21. Migliavacca, F., Petrini, L., Colombo, M.: Mechanical behaviour of coronary stents investigated through the finite element method. *Journal of Biomechanics* 35, 803–811 (2002)
22. Chua, S., Mc Donald, B., Hashmi, M.: Finite-element simulation of stent expansion. *Journal of Materials Processing Technology* 120, 335–340 (2002)

23. Zhu, H., Warner, J., Gehring, T., Friedman, M.: Comparison of coronary artery dynamics pre- and poststenting. *Journal of Biomechanics* 36, 689–697 (2003)
24. Benard, N., Coisne, D., Perrault, R.: Experimental study of laminar blood flow through an artery treated by a stent implantation. *Journal of Biomechanics* 36, 991–998 (2003)
25. Walke, W., Paszenda, Z.: Experimental and numerical biomechanical analysis of vascular stent. *Journal of Materials Processing Technology* 164–165, 1263–1268 (2005)
26. Sheth, S., Litvak, F., Fishbein, M., Forrester, J., Eigler, N.: Reduced thrombogenicity of polished and unpolished nitinolvs stainless steel slotted-tube stents in a pig coronary artery model. *Journal of the American College of Cardiology* 27, 197 (1997)
27. De Scheerder, I., Sohier, J., Wang, K.: Metallic surface treatment using electrochemical polishing decreases thrombogenicity and neointimal hyperplasia after coronary stent implantation in a porcine model. *European Heart Journal* 18, 153–156 (1997)
28. Gunn, J., Cumberland, D.: Stent coatings and local drug delivery. *European Heart Journal* 20, 1693–1700 (1999)
29. Verweire, I., Schacht, E., Qiang, B., Wang, K.: Evaluation of fluorinated polymers as coronary stent coating. *Journal of Materials Science: Materials in Medicine* 11, 207–212 (2000)
30. Bertrand, O., Sipehia, R., Mongrain, R., Rodes, J., Tardif, J.: Biocompatibility aspects of new stent technology. *Journal of the American College of Cardiology* 32, 562–571 (1998)
31. Christensen, K., Larsson, R., Elgue, G., Larsson, A.: Heparin coating of the stent graft – effects on platelets, coagulation and complement activation. *Biomaterials* 22, 349–355 (2001)
32. Michenatzis, G.: Comparison of haemocompatibility improvement of four polymeric biomaterials by two heparinization techniques. *Biomaterials* 24, 677–688 (2003)
33. Sousa, J., Morice, M., Serruys, P.: The RAVEL study – a randomized study with the sirolimus-coated BX Velocity balloon-expandable stent in the treatment of patients with de novo native coronary artery lesions. *The American Heart Association Scientific Sessions, Anaheim*, abstract 111305 (2001)
34. Grube, E., Silber, S., Hauptman, K.: Prospective, randomized, double-blind comparison of NIR stents coated with paclitaxel in a polymer carrier in de novo coronary lesions compared with uncoated controls. *The American Heart Association Scientific Sessions, Anaheim*, abstract 110945 (2001)
35. Paszenda, Z., Tyrlik-Held, J., Nawrat, Z., Żak, J., Wilczek, J.: Usefulness of passive-carbon layer for implants applied in interventional cardiology. *Journal of Materials Processing Technology* 157–158C, 399–404 (2004)
36. Paszenda, Z., Tyrlik-Held, J., Jurkiewicz, W.: Investigations of antithrombogenic properties of passive-carbon layer. *Journal of Achievements in Materials and Manufacturing Engineering* 17(1-2), 197–200 (2006)

Computer Enhanced Orthopedics

Wojciech Glinkowski^{1,2,3}

¹ Department of Anatomy, Center of Biostructure, Medical University of Warsaw

² Chair and Department of Orthopedics and Traumatology of Locomotor System,
Center of Excellence "TeleOrto" Medical University of Warsaw

³ Polish Telemedicine Society

Summary. The role of computers in orthopedic research and education and clinic is expanding rapidly. The computer assisted methods and modern technologies lately are changing musculoskeletal diagnostics, orthopedic surgery and rehabilitation. Technologies for computer assisted surgery (CAS) at the beginning were introduced into surgical practice for pre-operative planning and to enhance accuracy and safety for a variety of procedures. The introduction of highly demanding complex surgical procedures requires better visualization and detailed anatomy recognition intraoperatively. The new abilities to manipulate images during pre-operative planning increase an accuracy of surgical procedures. The orthopedic surgeon needs to be aware that technology driven methods are feasible and suitable nowadays. Computer-assisted methods for orthopedic surgery utilize the use of computers and robotic technology to assist in providing musculoskeletal care. Since its clinical implementation in neurosurgery, computer assisted methods in orthopedic surgery namely: surgical navigation, CAOS, CAD, distant learning, rapidly evolve in numerous applications. The final integration of all computerized applications creates new level of orthopedic surgical workflow of digital data that can be named Orthopedic PACS. Mentioned above methods have some clinically relevant implementations already, but further development is expected. The orthopedic surgeon should be aware of advantages as well pitfalls of its use. Clear understanding the goals, applications, and limitations of the computerized methods determine its successful current and future clinical use for the further improvement of the patients care. The future systems for daily practice should be characterized by easy learning, intuitive and friendly in use, and foolproof. The orthopaedic surgeon who understands and applies computerized technologies can expect further improvement in patients care.

1 Introduction

Computer technologies have become more and more integrated to orthopedic practice recently. Technologies for computer assisted surgery (CAS) were introduced into surgical practice for pre-operative planning and to enhance accuracy and safety for a variety of procedures. Advances in basic sciences, imaging, and advanced technology improve final clinical outcomes. The introduction of highly demanding complex surgical procedures requires better visualization and detailed anatomy recognition intraoperatively. The new abilities to manipulate images during pre-operative planning increase an accuracy of surgical treatment.

Recent development in musculoskeletal diagnostics, surgery and rehabilitation is partially due to introduction of modern technologies and interdisciplinary approach. The individualized patient care and management quality remains the main goals of modern, technology driven clinics in orthopedics. Clinical spectrum of computer enhancements begins with clinical teleeducation and passes through many clinical scenarios like: computer assisted diagnostics, preoperative planning, custom implant design, diagnostic screening, computerized navigation, robotics and computer assisted rehabilitation and telerehabilitation. Surgical digital workflow employing DICOM, databases, data warehouses; computer analytic tools, etc. create new way of dealing with patient's data. Described above new wave of orthopedic thinking creates Evidence Based Orthopedics that derives from computer assisted methods implementations. The definition of computer-assisted orthopedic surgery (CAOS) describes the use of computers and robotic technology to assist the orthopedic surgeon in providing clinical procedures. Computer-Assisted Orthopedic Surgery (CAOS) has made much progress over the last 10 years [1]. Computer assisted surgery (CAS) was initiated in neurosurgery and focused on pre-operative planning and increased intra-operative accuracy. The clinical applications of this technique expanded also in orthopaedic and trauma surgery, after its introduction in neurosurgery. When CAOS can be applied in the operating room only computer enhanced methods have wider spectrum equally preoperatively as postoperatively. Pre-operative planning precedes surgery by a pre-operative work-up, including planning and/or simulation. The 2D and 3D images, such as X-ray, CT- and MRI-scans are used for computing. Simulation can also play an important role in training and education of residents and surgeons who are not experienced yet. The surgeon can assess images during pre-operative planning and play with various operative solutions. In the operating room while using surgical navigation, the surgical instruments and the implants positioning are tracked, displayed and registered on a computer screen in relation to the patient's anatomy. The tracking sensor identifies and determines the instruments position during navigated surgery. The first application of such navigation system in orthopaedic and trauma surgery was for placement of lumbar transpedicular screws. The rationale for using navigation technology in pedicle screw placement was an incidence of incorrect placement of the screws under fluoroscopy guidance, ranging from 10-40% [2].

2 Computer Enhanced Orthopaedic Education

New methods from the "information age" progressively replace surgical teaching methods previously based on apprenticeship only. It is highly expected that surgeons will soon be able to acquire practical skills, theoretical knowledge as well validate and test their new competences from any location, using computer technology [3, 4]. Multimedia computer-aided learning in medicine will introduce important changes in surgical education [5].

A program consisting of computer-assisted instruction with interactive videodisc to teach residents in orthopaedic surgery the radiology of musculoskeletal injuries

could increase an interest of residents [6]. Radiography turned digital allows documenting interesting clinical findings with unprecedented ease, and storing them and creating large digital libraries of clinical results. Archiving, locating, and managing images, radiographs, and digital slide presentations are specific tasks for computer system use for orthopaedic surgeons nowadays. However, many surgical groups and practices are still not familiar with the computer technology available to initiate such systems [7]. A positive impact on the acquisition of musculoskeletal examination skills in medical students were found while utilizing of a specific computer-assisted learning (CAL) package the "Virtual Rheumatology" [8]. Wu et al. [8] have created "Virtual Orthopaedic European University" to support the implementation of online interactive courses for multimedia orthopaedic educational modules. They analyzed and developed case-based, problem-oriented learning, and different user-interaction scenarios and Learning Objects (LOs) system based on the pedagogical concept. They created an interactive course on Developmental Dislocation of Hip, as an example. The study by Thomas and Allen [10] suggested that computer-assisted education offers a reliable mode for teaching residents. They tested CD-ROM "Fundamentals of Orthopaedic Foot Care," produced by the American Academy of Orthopaedic Surgeons as a tool to provide nonoperative foot and ankle care education for orthopaedic and family practice residents. An interactive tutorial on anatomy, video demonstrations on selected topics in physical examination and basic treatment of nonoperative problems of the foot and ankle, and patient education information sheets on multiple common foot disorders were stored in educational format CD-ROM. That program was assumed as reliable in an orthopaedic residency program to achieve ACGME required competency of "medical knowledge" in evaluation and nonoperative management of common foot and ankle problems. Orthopedic resources on the Internet (Orthogate, OrthoNet, American Academy of Orthopaedic Surgeons site, Orthopedic Hyperguide, WorldOrtho, Wheelless's Textbook of Orthopaedics, Orthoteers, AO North America site, University of Iowa Virtual Hospital texts and South Australian Orthopaedic Registrars' Notebook) are becoming an integral part of orthopaedic education every day [11]. Wheelless's Textbook of Orthopaedics [12], the American Academy of Orthopaedics Surgeons website [13], and Orthopedics Hyperguide [14] are commonly used online sites by the Orthopaedic residents when looking for clinical information and online practice examinations when preparing for the in-training (OITE) or board examinations. Attending physicians do not use the Internet resources so often but their interest toward Internet updated orthopedic knowledge is rising up. Physicians can have access to expert's lectures and discussions through internet. Flash enhanced videoconference and chat were successfully used in our study [15]. The telementored surgeon can be instructed also by the expert specialist during the procedure through videoconferencing [16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27]. Teleeducation, teleteaching, teletraining, telementoring have been introduced into daily medical educational services to lifelong stay updated to medical sciences progress for health-care professionals. Information and communication technologies (ICTs) play an important role in the delivery of distance learning. Videoconferencing allows successful development of the concepts of medical and surgical

teleeducation [28]. Internet-based technologies (e.g. email and the World Wide Web) and videoconferencing become the most common teleeducation technologies being used [29, 30]. High-resolution video imaging cameras can transmit accurate visualization of the surgical field and share the surgical procedure with a remote audience with visual transmission of the surgical field, which is otherwise very difficult to share surgery [24]. Allen et al. [31] utilized videoconferencing to provide distance education for medical students, physicians and other health-care professionals, such as nurses, physiotherapists and pharmacists. They found that videoconferencing has been well accepted by faculty staff and by learners, as it enables them to provide and receive CME without traveling long distances. Videoconferencing plays very important role in international medical scientific linking [32]. Telemedicine experience is able to improve greatly the quality of care available to travelers and migrant workers in cases where the patient cannot communicate with the attending physician because of a language barrier. Ricci et al. [33] allows providing the continuing medical education for physicians in rural communities who have limited access to CME opportunities. The study conducted by Krupinski et al. [34] has supported the thesis that tele-education broadcasts may supplement many CME activities but may not replace them. University educational videoconferencing technology has the potential for bridging urban-rural divide for the benefit of new and continuing health professions education [35]. The internet as an international communication network between international centers of excellence allows multidisciplinary exchange of medical information among specialists from different countries and cultures. Improvements in videoconferencing systems permit interactive distant communication and decrease their costs. Own studies confirmed that distance education can be provided for medical students, physicians and other healthcare professionals, such as nurses, physiotherapists and pharmacists utilizing telemedicine videoconferences [15]. Videoconferencing acceptance seems to be high for participants as well faculty members. Teleeducation that enables providing and receiving CME without traveling long distances was widely appreciated. Another way to supplement teaching targets in building searchable radiology museums based on PACS and cataloguing a whole DICOM datasets [36].

3 Orthopedic PACS

In general a PACS system can simplify and speed up an IT integrated hospital workflow. A well-planned, fully integrated digital Orthopedic Department can simplify workflow throughout the hospital. PACS technology, empowered with speech recognition systems and web distribution tools, can deliver significant benefits to a health care institution [37, 38, 39, 40, 41, 42]. The imaging procedures workflow may influence the length of hospital stay for inpatients requiring imaging procedures and demonstrates how the PACS introduction leads to benefits from eliminating transmission time. MRI, CT, ultrasound and radiography exams are delivered to various departments in the same hospital, including orthopedics. Orthopaedic Department can be positively benefited by

PACS. Nitrosi et al. [43] noted reasonable savings per patient bed/days. Financial criteria should not be the only driving factor in choosing to adopt a fully integrated PACS solution. The workload of the imaging department, particularly servicing for orthopedic emergencies produce significant expenses for the hospital and digital imaging technologies are able to cut these expenses. Orthopedic surgeons strongly rely on image display in the OR to help determine their operative approach and to guide their surgery. Orthopedic PACS (Ortho-PACS) offers unique intraoperative imaging capabilities including rapid image retrieval and improved archival, cine review, the ability to modify image contrast, and the ability to obtain direct quantitative measurements of the musculoskeletal structures. The increased accessibility and availability of images throughout the hospital enables improvement in time management and in patient care. Korb et al. [44] clarified definition of an S-PACS system. A Surgical Picture Archiving and Communication System (S-PACS) design is focused on better integration of surgical imaging assistance systems inside the operating room environment. The communication and processing platform need to link together hospital information system, radiology information system and surgical information system that derive additionally from intraoperative data. Intraoperative imaging is used daily in orthopedic surgery practice. Reiner et al. [45] have determined the acceptance and clinical utility of a large scale picture archiving and communication system (PACS) for vascular surgery. Pomerantz et al. [46] suggested that the design of a hospital-wide PAC System must have the flexibility to accommodate the specific requirements of a wide variety of end-users in their unique hospital environments. They analyzed of the efficacy of a picture archiving and communication system (PACS) in the surgical domain. Direct imaging observation in the operating room (OR) and surgical outpatient clinic is determined by patterns by routine clinical PACS use. Specific modifications in PACS design for surgical use requires some improvement the currently distributed PACS particularly for orthopaedics. PACS-based Arthroscopy Image Acquisition Workstation requires a Multimedia Video Card. The arthroscopy video is digitized and captured dynamically or statically into computer. A variety of functions such as the arthroscopy video's acquisition and display, editing, processing, managing, storage, printing and communication of related information flows through the workstation and link to hospital PACS. It can also act as an independent arthroscopy diagnostic system.

Watkins et al. [47] examined the influence of a PACS on the length of stay for patients receiving total hip replacement (THR) or total knee replacement (TKR) procedures. They have shown an apparent average length of hospital stay reduction of 25% for total knee replacement patients but no similar reduction for THR patients. Their conclusions could not be explained by PACS system introduction. PACS workstation using computer-assisted measurement software can be used for computer-assisted angle measurement on digital total-leg radiographs with high reliability [48]. The pixel density value of PACS was found as a rapid and easy method for the detection of bone density changes in distraction osteogenesis [49].

4 Computer Assisted Preoperative Planning

Computer technologies were introduced into orthopedic practice for pre-operative planning and to enhance accuracy and safety for a variety of procedures. PACS and computed plain radiography are used in preoperative planning [50]. Preoperative planning may determine anatomic references for size and implant orientation prior the surgery [51]. The majority of orthopaedic surgery planning programs aim to achieve graphical representation of a patient's anatomy. The planning task focuses on the accuracy with which each user was able to position the implant. The type of visualization modality used to represent CT data on final accuracy of the planning operation [52]. Simulations can help orthopaedic surgeons to develop better preoperative plans. Surgical simulations are particularly appropriate for the large volume and expense of joint replacement procedures in orthopaedics. The model of the implantation procedure for cementless acetabular and femoral components in total hip replacement surgery can be simulated. The simulator based upon finite element analysis may predict the early postoperative mechanical environment that results from a proposed surgery [53].

Computer-assisted preoperative planning (CAPP) is a complex approach to plan surgical treatment. It includes computer-assisted design (CAD) principles, expert knowledge utilization and facilities for decision making process support. CAPP application by Zdravkovic and Bilic [54] was designed to preoperatively plan a corrective osteotomy of a malunited fracture at the distal end of the radius. Component placement critically affects the performance and longevity of total hip replacements (THRs) [55, 56]. Because of limitations of observation and anatomic orientation imposed by the operative site, selection of the correct size, and position of the acetabular and femoral components is best done through preoperative planning. Currently, two-dimensional templates of prosthetic components with clinical radiographs are used for preoperative planning; however, this method has the inherent limitation of plain radiographs only. Computerized approach allows surgeons to template with superior accuracy. Observed technologic advances in surgical technique, computer-based preoperative planning tools should prove all the more essential to reliable component placement. Viceconti et al. [57] compared accuracy and the repeatability in planning total hip replacements with the conventional templates on radiographs to that attainable on the same clinical cases when using CT-based planning software. The sizes of the cementless components planned with new computer aided preoperative planning system called Hip-Op and with standard templates were compared to those effectively implanted. They observed comparable repeatability of the Hip-Op system to that of the template Procedure. The repeatability of the preoperative planning with the Hip-Op system was independent from surgeon experience. Their study has clearly shown the advantages (accuracy and repeatability) of a three-dimensional computer-based preoperative planning over the traditional template planning, particularly for anatomically deformed cases. The surgical planning was advantageous especially for the socket and for less experienced surgeons.

5 Computer Assisted Surgical Navigation in Orthopedic Surgery

Modern navigation technology appears to be acquiring an established place in the orthopedic surgery for helping the surgeon to apply his manual skills with greater precision and effectiveness. Computer assisted surgical navigation in orthopedic surgery describes navigation systems that provide additional information during a procedure integrating preoperative planning with intraoperative execution [58]. In orthopedic trauma computer assisted surgery has been most frequently mentioned as an adjunct to pelvic, acetabular or femoral fractures [58, 59, 60]. Navigation is applied while percutaneous ilio-sacral screws, percutaneous fixation of hip fractures, alignment and fixation of long bone, and spinal fractures are performed. Navigation has become increasingly integrated into orthopaedic surgery, especially in the area of endoprosthetic procedures, including minimally invasive approach [61]. Surgical navigation is considered as novel approach in arthroplasties. In total hip arthroplasty the acetabular component orientation, placement of femoral component and in total knee replacement alignment of the femoral and tibial components enhanced by navigation influence on long-term outcomes. Diminishing of frequency of arthroplasty components malalignment utilizing navigation and computer assisted surgery methods results in decrease of instability and reoperation [60, 62]. Determination of the amount of variation in conventional acetabular cup positioning (radiological inclination and anteversion) in view of different factors that could influence the measured angles was checked radiologically in 950 patients who received a cementless total hip replacement [63]. The authors found that application of computer-aided navigation in the placement of acetabular cups was able to improve accuracy and reproducibility considerably in total hip arthroplasty. The qualifications of the surgeon, the implanted model, as well as the operated side did not have a significant influence on their results. Conventional surgery uses C-arm fluoroscopy guiding to position the implant in one plane and then get additional images in other planes in a trial and error fashion to ensure that the device has been properly placed. Computed assisted surgery adds new values to the existing image guidance using C-arm fluoroscopy resulting in more precise interlocking, placement, and radiation exposure time as outcomes. CAS navigation involves three steps; data acquisition, registration and tracking. Data acquisition is based on image based (C-arm fluoroscopy, CT/MRI) or imageless systems. Anatomical landmarks centers of rotation of the hip, knee or ankle, or visual are used in the imageless systems. Related images and anatomical position in the surgical field are referred by registration. Infrared markers are surgically placed in imageless systems or surface matching technique can be used to match shapes of the bone surface model generated from preoperative images to surface data points collected during surgery. Markers, sensors and measurement devices providing feedback during surgery regarding the orientation and relative position of bone anatomy are used for tracking. Better target orientation allows achieving improved final alignment of the prosthesis [64, 65]. Simplification of the instrumentation along with the use of imageless systems has increased the ease of use for the orthopaedic surgeon. In image free navigated knee arthroplasty the hip

centre is determined by rotary motion of the femur (pivoting). Actual algorithms used for the registration process of the lower extremity mechanical axis and the articular surfaces reveal valid and reproducible results [66]. The accuracy of this functional hip centre determination in vivo during pivoting is maximal for hip centre determination of 20 to 30 degrees in the sagittal plane and 30 to 40 degrees in the frontal plane [62]. Matziolis et al. [67] confirmed that arthritis of the hip is not a contraindication for functional determination of the hip centre. With the help of navigation, it is possible to achieve a higher degree of precision in total hip and knee implant placement, high tibial osteotomy, or more precise correction of the axis at the osteotomy of the distal radius and finally improved procedure reproducibility. Large prospective clinical studies comparing conventional techniques versus computer assisted navigation are only available for total knee arthroplasty. One of the major goals of total knee replacement is to create a stable knee while recreating the patient's mechanical axis to restore limb alignment. Using the OrthoPilot[®] system the process of final implant placement can be enhanced [61]. The navigated prosthesis placement can truly extend the longevity of an implant requires continued, long-term observation. Despite insufficient strong scientific evidence to allow conclusions regarding the superiority computer assisted technologies for orthopedic surgery compared to conventional methods in long-term effectiveness, short term outcomes are rather convincing. The navigation systems show that the technology can be adapted to the requirements of daily surgical practice, without compromising its utility for the surgeon and the patient.

6 Robotics in Orthopedic Surgery

The semi-autonomous procedure of robotic surgery by a robotic arm under direct or indirect physician control was performed first time human surgery in 1992 in California using the ROBODOC[®] system [1]. The field of medical robotics is still emerging. Clinical applications utilizing robots have been applied in many specialties including orthopedics [68]. Despite the potential to improve the precision and capabilities of physicians, the number of robots in clinical use is still very small. Orthopedic surgery is considered as ideal specialty for the application of robotic systems [69]. Improved accuracy, precision in the preparation of bone surfaces, more reliable and reproducible outcomes, and greater spatial accuracy are the main advantages of robot-assisted orthopedic surgery over conventional orthopedic techniques. Total hip and knee replacement, tunnel placement for reconstruction of knee ligaments, orthopedic trauma and spinal procedures utilizing robot-assisted orthopedic surgery applications currently are investigated [1, 68, 69, 70]. Several short-term studies demonstrate the feasibility of robotic applications in orthopedics. Commercially available robotic systems can be classified as passive or active devices. Robot-assisted orthopedic procedures can be categorized as positioning or milling/cutting. Levels of accuracy, precision, and safety of robot-assisted orthopedic surgery can not be achieved with computer assisted orthopedic surgery. The development of robot-assisted orthopedic surgery is expensive. The immature

and not wide spread robotic orthopedics has high potential to transform the way of orthopedic procedures in the future.

7 Virtual Reality Orthopedic Surgery

The orthopedic surgical experience grows with special concern in omitting surgical errors usually as apprenticeship. That way is time consuming. The developing virtual reality simulation allows training for orthopaedic residents in various procedures in surgery before live-patient operating room experience. Surgical training systems using virtual reality simulation techniques offer a cost-effective alternative to traditional training methods. It is a promising medical area for 3-D computer graphics and virtual reality techniques and applications. Using combined 3-D reconstruction and virtual environment (VE) technologies to train clinicians and to help surgeons plan patient-specific, surgery for trauma from accidents and reconstruction surgery [71]. Virtual reality system may be especially useful for difficult surgical procedures performed in orthopedic and reconstructive surgery departments. The system can be used to teach interns, train residents and visiting staff. Various developments have appeared lately in segmentation, personal-computer-based real-time volume visualization, tissue analysis with topological change in real-time using finite element analysis, and simulation of soft tissue cutting with tactile feedback.

The simulators appear as specialized tools for surgeons to simulate a use various surgical instruments to operate on virtual rigid anatomic structures, prostheses and bone grafts. The simulation may be applied to every procedure on the rigid structures for complex orthopedic surgeries, including arthroplasty, arthroscopy, osteotomy, open reduction of fractures and amputation [71, 72, 73, 74] and fluoroscopic navigation [75]. An arthroscopic virtual reality realistic human knee anatomy simulator for training orthopaedic residents in arthroscopic surgery has been developed by Cannon [76]. Incorporated active force-feedback haptic technology for virtual education program was implemented to reach a proficiency standard in the techniques and protocol for an arthroscopic knee examination.

Hsieh et al. [77] described the 3D virtual reality simulation system to provide preoperative simulation to verify that the osteotomy and fusion procedures chosen to treat musculoskeletal defects are appropriate. Similarly, augmented reality systems begin to be used everyday in medical training, preoperative planning, preoperative and intraoperative data visualization, and intraoperative tool guidance [78]. Augmented reality is a display technique that combines supplemental information with the real world environment.

8 Computerized Design of Orthopedic Implants

The custom orthopedic implant design is considered in selected cases because it is expected that improved implant fit may increase the longevity of total joint

replacements. Computed tomography and interactive image processing methods are used to generate the individual implant design [79, 80, 81]. Computer modeling techniques, computer graphics and software are implemented to computer aided orthopedic design in creating a custom implant with good anatomic conformity. The CAD/CAM techniques are usually used in the production of custom-made hip replacements utilizing digitized radiographic data [81]. The development of custom endoprostheses has target into a niche in the joint replacement market. The theoretical model are created to test the prosthesis/bone geometry, interface or remodeling for an individual patient [82]. Reconstruction of the bone morphology remains a fundamental step in prosthesis design [83, 84, 85].

9 Telemedicine in Orthopedics

Telemedicine widely enters into all specialties of modern medicine including orthopedics and musculoskeletal traumatology [86]. Teleconsultations in the polytrauma cases may decrease hospitalization rate, complication rates and risk. Synchronous teleconsultations can be scheduled online. More often utilized are asynchronous consultations [87, 88]. Second opinions in orthopaedics seem to become an additional teleconsulting way of telemedicine service [89]. The easy integration into clinical practice can be granted to the radiology images (radiographs, computerized tomography scans, magnetic resonance imaging scans and ultrasound scans), video-messages and still images. That information can be transmitted as the asynchronous consultations. The diagnostic quality of the information transmitted still remains questionable unless it is DICOM file. Ricci and Borrelli [90] have attempted to determine whether teleradiology improved clinical decision making for the treatment of patients with acute fractures based on Level 1 trauma center communication to the attending orthopedic surgeon on call. In each case, an orthopaedic junior resident performed the emergency department consultation utilizing digitized and electronically transmitted radiographs to the attending specialist. They have confirmed that routine use of electronically transmitted digitized radiographic images has the potential to improve clinical decision making for the care of patients with acute fractures. Teleconsultations save time and prevent the unnecessary transfer of patients to main hospitals [91]. Teleradiology is defined as an electronically transmitting radiographic image files from one location to another [92]. Technologic advances in digital imaging, telecommunications, digital storage, and viewing technologies have made teleradiology readily available and reasonably affordable. The teleradiology system consists of a sending station, a transmission network, a storage device, a viewing station and, a software package. Teleradiology has been shown to improve diagnostic accuracy, disposition planning of patients from emergency departments or outlying hospitals, and planning of surgical procedures due to digital images transfer. Consultations via videoconferencing and traditional outpatient clinic visits present no particular difference between groups. The videoconferencing has been confirmed as a valid alternative to outpatient clinic visits for orthopaedic specialist consultations.

10 Conclusions and Final Remarks

Computer technology driven methods are feasible and suitable for the orthopaedic surgeon. Computer enhanced methods are used recently in orthopedic surgery namely: Computer enhanced orthopaedic education, Orthopedic PACS, Computer assisted preoperative planning, Computer assisted surgical navigation in orthopedic surgery, Robotics in orthopedic surgery, Virtual reality orthopedic surgery, Computerized design of orthopedic implants and Telemedicine. These methods have already some clinically relevant implementations. Computer augmented orthopaedic surgery improves standard procedures by pre-operative planning, intraoperative navigation, smart tools and remote surgery technologies. The advantages of computer assisted methods include geometric precision, reproducibility, perfect "memory," lack of fatigue, and insensitivity to radiation. Although initial experiences appear promising, most of methods are still complex, sensitive to failures, expensive and long with learning curve. The orthopedic surgeon should be aware of pitfalls that may appear in registration, tracking, instability of software and others. Clear understanding the goals, applications, and limitations of the computerized methods determine its successful and proper clinical use. The basic principles of managing radiographic and anatomic data, joint arthroplasty, deformity correction, and spinal and trauma surgery remain still crucial for utilizing computer enhanced methods. The future computer assisted systems for daily practice must become easier to learn and intuitive in use, as well user friendly and foolproof. The orthopaedic surgeon who understands and applies computerized technologies needs to gain the further improvement of the patients care.

References

1. Specht, L.M., Koval, K.J.: Robotics and Computer-Assisted Orthopaedic Surgery. *Bull. Hosp. J. Dis.* 60, 3–4 (2001–2002)
2. Merloz, P., Tonetti, J., Pittet, L., Coulomb, M., Lavallee, S., Sautot, P.: Pedicle screw placement using image-guided techniques. *Clin. Orthop.* 354, 39–48 (1998)
3. Demartines, N., Mutter, D., Vix, M., Leroy, J., Glatz, D., Rosel, F., Harder, F., Marescaux, J.: Assessment of telemedicine in surgical education and patient care. *Ann. Surg.* 231(2), 282–291 (2000)
4. Malassagne, B., Mutter, D., Leroy, J., Smith, M., Soler, L., Marescaux, J.: Teleducation in Surgery: European Institute for TeleSurgery Experience. *World J. Surg.* 25, 1490–1494 (2001)
5. Grosfeld, J.L.: Presidential Address. Visions: medical education and surgical training in evolution. *Arch. Surg.* 134(6), 590–598 (1999)
6. Chew, F.S., Smirniotopoulos, J.G.: Teaching skeletal radiology with use of computer-assisted instruction with interactive videodisc. *J. Bone Joint Surg. Am.* 77(7), 1080–1086 (1995)
7. Gomoll, A.H., Thornhill, T.S.: Image catalogs. *Clin. Orthop. Relat. Res.* (421), 29–34 (2004)

8. Vivekananda-Schmidt, P., Lewis, M., Hassell, A.B.: Cluster randomized controlled trial of the impact of a computer-assisted learning package on the learning of musculoskeletal examination skills by undergraduate medical students. *Arthritis Rheum.* 15, 53(5), 764–771 (2005)
9. Wu, T., Zimolong, A., Schiffers, N., Ohnsorge, J.A., Radermacher, K.: Developing authoring tools for web-based multi-media orthopedics education modules. *Biomed. Tech (Berl)* 47(suppl. 1 pt 1), 350–353 (2002)
10. Thomas, R.L., Allen, R.M.: Use of computer-assisted learning module to achieve ACGME competencies in orthopaedic foot and ankle surgery. *Foot Ankle. Int.* 24(12), 938–941 (2003)
11. Sinkov, V.A., Andres, B.M., Wheelless, C.R., Frassica, F.J.: Internet-based learning. *Clin. Orthop. Relat. Res.* 421, 99–106 (2004)
12. <http://www.wheelsonline.com/>
13. <http://www.aaos.org/>
14. <http://www.ortho.hyperguides.com/>
15. Glinkowski, G., Małosa, K., Pawlica, S., Marasek, K., Górecki, A.: Medical interactive teleeducation via Internet based videoconferencing. In: Piętka, E., Łęski, J., Franiel, S. (eds.) *Proceedings of the XI International Conference Medical Informatics & Technology*, pp. 254–258. MIT Press, Cambridge (2006), <http://www.itib.edu.pl/mit/papers/index.htm>
16. Byrne, J.P., Mughal, M.M.: Telementoring as an adjunct to training and competence-based assessment in laparoscopic cholecystectomy. *Surg. Endosc.* 14(12), 1159–1161 (2000)
17. Byrne, J.P., Mughal, M.M.: Telementoring in laparoscopic cholecystectomy: a useful adjunct in training and assessment of higher surgical trainees. *Br. J. Surg.* 87, 362–373 (2000)
18. Gandsas, A., McIntire, K., George, I.M., Witzke, W., Hoskins, J.D., Park, A.: Wireless live streaming video of laparoscopic surgery: a bandwidth analysis for handheld computers. *Stud Health Technol. Inform.* 85, 150–154 (2002)
19. Gandsas, A., McIntire, K., Montgomery, K., Bumgardner, C., Rice, L.: The personal digital assistant (PDA) as a tool for telementoring endoscopic procedures. *Stud Health Technol. Inform.* 98, 99–103 (2004)
20. Gul, Y.A., Wan, A.C., Darzi, A.: Undergraduate surgical teaching utilizing telemedicine. *Med. Educ.* 33(8), 569–596 (1999)
21. Latifi, R., Peck, K., Satava, R., Anvari, M.: Telepresence and telementoring in surgery. *Stud Health Technol. Inform.* 104, 200–206 (2004)
22. Mendez, I., Hill, R., Clarke, D., Kolyvas, G.: Robotic long-distance telementoring in neurosurgery. *Neurosurgery* 56(3), 434–440 (2005)
23. Pradeep, P.V., Mishra, S.K., Vaidyanathan, S., Nair, C.G., Ramalingam, K., Basnet, R.: Telementoring in endocrine surgery: preliminary Indian experience. *Teled. J. E. Health* 12(1), 73–77 (2006)
24. Rafiq, A., Moore, J.A., Zhao, X., Doarn, C.R., Merrell, R.C.: Digital Video Capture and Synchronous Consultation in Open Surgery. *Ann. Surg.* 239(4), 567–573 (2004)
25. Schlag, P.M., Moesta, K.T., Rakovsky, S., Grasczew, G.: Telemedicine: the new must for surgery. *Arch. Surg.* 134, 1216 (1999)
26. Seabajang, H., Trudeau, P., Dougall, A., Hegge, S., McKinley, C., Anvari, M.: Telementoring: an important enabling tool for the community surgeon. *Surg. Innov.* 12(4), 327–331 (2005)
27. Seabajang, H., Trudeau, P., Dougall, A., Hegge, S., McKinley, C., Anvari, M.: The role of telementoring and telerobotic assistance in the provision of laparoscopic colorectal surgery in rural areas. *Surg. Endosc.* 20(9), 1389–1393 (2006)

28. Neame, R., Murphy, B., Stitt, F., Rake, M.: Universities without walls: evolving paradigms in medical education. *B.M.J.* 319, 1296 (1999)
29. Curran, V.R.: Tele-education. *J. Telemed. Telecare* 12(2), 57–64 (2006)
30. Glinkowski, W., Bogdan, C.: WWW-Based e-Teaching of Normal Anatomy as an Introduction to the Telemedicine and e-Health. *Telemed J. E. Health* 13(5), 49–58 (2007)
31. Allen, M., Sargeant, J., MacDougall, E., Proctor-Simms, M.: Videoconferencing for continuing medical education: from pilot project to sustained programme. *J. Telemed. Telecare* 8(3), 131–137 (2002)
32. Anogianakis, G., Ilonidis, G., Anogeianaki, A., Milliaras, S., Klisarova, A., Temelkov, T., Milliaras, V.E.: A clinical and educational telemedicine link between Bulgaria and Greece. *J. Telemed. Telecare* 9(suppl. 2), 2–4 (2003)
33. Ricci, M.A., Caputo, M.P., Callas, P.W., Gagne, M.: The use of telemedicine for delivering continuing medical education in rural communities. *Telemed. J. E. Health* 11(2), 124–129 (2005)
34. Krupinski, E.A., Lopez, A.M., Lyman, T., Barker, G., Weinstein, R.S.: Continuing education via telemedicine: analysis of reasons for attending or not attending. *Telemed. J. E. Health* 10(3), 403–409 (2004)
35. Cunningham, B.J., Stamm, B.H.: The education part of telehealth. *Rural Remote Health* 5(4), 400 (2005)
36. Toms, A.P., Kasmai, B., Williams, S., Wilson, P.: Building an anonymized catalogued radiology museum in PACS: a feasibility study. *Br. J. Radiol.* 79(944), 666–671 (2006)
37. Ondo, K.: PACS direct experiences: Implementation, selection, benefits realized. *J. Digit. Imaging* 17(4), 249–252 (2004)
38. Huang, H.K.: *PACS and Imaging Informatics: Basic Principles and Applications*. Wiley & Sons, New York (2004)
39. Piętka, E., Pośpiech-Kurkowska, S., Gertych, A., Cao, F.: Integration of computer assisted bone age assessment with clinical PACS. *Comput. Med. Imaging Graph* 27(2-3), 217–228 (2003)
40. Cao, F., Huang, H.K., Piętka, E., Gilsanz, V.: Digital hand atlas and web-based bone age assessment: system design and implementation. *Comput. Med. Imaging Graph* 24(5), 297–307 (2000)
41. Huang, H.K., Wong, S.T., Piętka, E.: Medical image informatics infrastructure design and applications. *Med. Inform (Lond)* 22(4), 279–289 (1997)
42. McNitt-Gray, M.F., Piętka, E., Huang, H.K.: Image preprocessing for a picture archiving and communication system. *Invest. Radiol.* 27(7), 529–535 (1992)
43. Nitrosi, A., Borasi, G., Nicoli, F., Modigliani, G., Botti, A., Bertolini, N.P.: A Filmless Radiology Department in a Full Digital Regional Hospital: Quantitative Evaluation of the Increased Quality and Efficiency. *J. Digital Imaging* 20(2), 140–148 (2007)
44. Korb, W., Bohn, S., Burgert, O., Dietz, A., Jacobs, S., Falk, V., Meixensberger, J., Strauss, G., Trantakis, C., Lemke, H.U.: Surgical PACS for the digital operating room. Systems engineering and specification of user requirements. *Stud Health Technol. Inform.* 119, 267–272 (2006)
45. Reiner, B.I., Siegel, E.L., Hooper, F., Pomerantz, S.M., Protopapas, Z., Pickar, E., Killewich, L.: Picture archiving and communication systems and vascular surgery: clinical impressions and suggestions for improvement. *J. Digit. Imaging* 9(4), 167–171 (1996)

46. Pomerantz, S.M., Siegel, E.L., Protopapas, Z., Reiner, B.I., Pickar, E.R.: Experience and design recommendations for picture archiving and communication systems in the surgical setting. *J. Digit Imaging* 9(3), 123–130 (1996)
47. Watkins, J.R., Bryan, S., Muris, N.M., Buxton, M.J.: Examining the influence of picture archiving communication systems and other factors upon the length of stay for patients with total hip and total knee replacements. *Int. J. Technol. Assess. Health Care* 15(3), 497–505 (1999)
48. Sailer, J., Scharitzer, M., Peloschek, P., Giurea, A., Imhof, H., Grampp, S.: Quantification of axial alignment of the lower extremity on conventional and digital total leg radiographs. *Eur. Radiol.* 15(1), 170–173 (2005)
49. Shim, J.S., Chung, K.H., Ahn, J.M.: Value of measuring bone density serial changes on a picture archiving and communication systems (PACS) monitor in distraction osteogenesis. *Orthopedics* 25(11), 1269–1272 (2002)
50. Parikh, S.N., Brody, A.S., Crawford, A.H.: Use of a picture archiving and communications system (PACS) and computed plain radiography in preoperative planning. *Am. J. Orthop.* 33(2), 62–64 (2004)
51. Tannast, M., Langlotz, U., Siebenrock, K.A., Wiese, M., Bernsmann, K., Langlotz, F.: Anatomic referencing of cup orientation in total hip arthroplasty. *Clin. Orthop. Relat. Res.* 436, 144–150 (2005)
52. Viceconti, M., Lattanzi, R., Zannoni, C., Cappello, A.: Effect of display modality on spatial accuracy of orthopaedic surgery pre-operative planning applications. *Med. Inform. Internet Med.* 27(1), 21–32 (2002)
53. O’Toole III, R.V., Jaramaz, B., DiGioia III, A.M., Visnic, C.D., Reid, R.H.: Biomechanics for preoperative planning and surgical simulations in orthopaedics. *Comput. Biol. Med.* 25(2), 183–191 (1995)
54. Zdravkovic, V., Bilic, R.: Computer-assisted preoperative planning (CAPP) in orthopaedic surgery. *Comput. Methods Programs Biomed.* 32(2), 141–146 (1990)
55. Sugano, N., Ohzono, K., Nishii, T., Haraguchi, K., Sakai, T., Ochi, T.: Computed-tomography-based computer preoperative planning for total hip arthroplasty. *Comput. Aided Surg.* 3(6), 320–324 (1998)
56. Noble, P.C., Sugano, N., Johnston, J.D., Thompson, M.T., Conditt, M.A., Engh Sr., C.A., Mathis, K.B.: Computer simulation: how can it help the surgeon optimize implant position? *Clin Orthop. Relat. Res.* 417, 242–252 (2003)
57. Viceconti, M., Lattanzi, R., Antonietti, B., Paderni, S., Olmi, R., Sudanese, A., Toni, A.: CT-based surgical planning software improves the accuracy of total hip replacement preoperative planning. *Med. Eng. Phys.* 25(5), 371–377 (2003)
58. Schep, N.W.L., Broeders, I.A.M.J., van der Chr, W.: Computer assisted orthopaedic and trauma surgery. *Injury* 4, 299–306 (2003)
59. Slomczykowski, M.A., Hofstetter, R., Sati, M., et al.: Novel computer-assisted fluoroscopy system for intraoperative guidance: feasibility study for distal locking of femoral nails. *J. Orthop. Trauma.* 15(2), 122–131 (2001)
60. Leenders, T., Vandeveld, D., Mahieu, G., Nuyts, R.: Reduction in variability of acetabular cup abduction using computer assisted surgery: a prospective and randomized study. *Computer Aided Surg.* 7(2), 99–106 (2002)
61. Uchowicz, M., Górecki, A., Purski, K., Jabłoński, T.: Minimally invasive techniques in total hip replacement—our clinical experience. *Ortopedia, traumatologia, rehabilitacja* 9(1), 15–24 (2007)
62. Saragaglia, D., Picard, F., Chaussard, C., et al.: Computer-assisted knee arthroplasty: comparison with a conventional procedure. Results of 50 cases in a prospective randomized study. *Rev. Chir. Orthop. Reparatrice Appar. Mot.* 87(1), 18–28 (2001)

63. Leichtle, U., Gosselke, N., Wirth, C.J., Rudert, M.: Radiologic evaluation of cup placement variation in conventional total hip arthroplasty. *Rofo*, vol. 179(1), pp. 46–52 (2007)
64. Suhm, N., Jacob, A.L., Nolte, L.P., et al.: Surgical navigation based on fluoroscopy-clinical application for computer-assisted distal locking of intramedullary implants. *Comput. Aided Surg.* 5(6), 391–400 (2000)
65. Jenny, J.Y., Boeri, C.: Computer-assisted implantation of total knee prostheses: a case-control comparative study with classical instrumentation. *Comput. Aided Surg.* 6(4), 217–220 (2001)
66. Hufner, T., Kendoff, D., Citak, M., Geerling, J., Krettek, C.: Precision in orthopaedic computer navigation. *Orthopade* 35(10), 1043–1055 (2006)
67. Matziolis, G., Krockner, D., Tohtz, S., Weiss, U., Perka, C.: Accuracy of determination of the hip centre in navigated total knee. *Arthroplasty Z. Orthop. Ihre Grenzgeb* 144(4), 362–366 (2006)
68. Cleary, K., Nguyen, C.: State of the art in surgical robotics: clinical applications and technology challenges. *Comput. Aided Surg.* 6(6), 312–328 (2001)
69. Adili, A.: Robot-assisted orthopedic surgery. *Semin Laparosc Surg.* 11(2), 89–98 (2004)
70. Mendez, I., Hill, R., Clarke, D., Kolyvas, G.: Robotic long-distance telementoring in neurosurgery. *Neurosurgery* 56(3), 434–440 (2005)
71. Heng, P.A., Cheng, C.Y., Wong, T.T., Wu, W., Xu, Y., Xie, Y., Chui, Y.P., Chan, K.M., Leung, K.S.: Virtual reality techniques. Application to anatomic visualization and orthopaedics training. *Clin Orthop. Relat. Res.* 442, 5–12 (2006)
72. Mabrey, J.D., Cannon, W.D., Gillogly, S.D., Kasser, J.R., Sweeney, H.J., Zarins, B., Mevis, H., Garrett, W.E., Poss, R.: Development of a virtual reality arthroscopic knee simulator. *Stud Health Technol. Inform.* 70, 192–194 (2000)
73. Tsai, M.D., Hsieh, M.S., Jou, S.B.: Virtual reality orthopedic surgery simulator. *Comput. Biol. Med.* 31(5), 333–351 (2001)
74. Poss, R., Mabrey, J.D., Gillogly, S.D., Kasser, J.R., Sweeney, H.J., Zarins, B., Garrett, W.E., Cannon Jr., W.D.: Development of a virtual reality arthroscopic knee simulator. *J. Bone Joint Surg. Am.* 82-A(10), 1495–1499 (2000)
75. Jaramaz, B., Eckman, K.: Virtual reality simulation of fluoroscopic navigation. *Clin Orthop. Relat. Res.* 442, 30–34 (2006)
76. Cannon, W.D., Eckhoff, D.G., Garrett, W.E., Hunter, R.E., Sweeney, H.J.: Report of a group developing a virtual reality simulator for arthroscopic surgery of the knee joint. *Clin. Orthop. Relat. Res.* 442, 21–29 (2006)
77. Hsieh, M.S., Tsai, M.D., Chang, W.C.: Virtual reality simulator for osteotomy and fusion involving the musculoskeletal system. *Comput. Med. Imaging. Graph* 26(2), 91–101 (2002)
78. Blackwell, M., Morgan, F., DiGioia, A.M.: Augmented reality and its future in orthopaedics. *Clin. Orthop. Relat. Res.* (354), 111–122 (1998)
79. Bechtold, J.E.: Application of computer graphics in the design of custom orthopedic implants. *Orthop. Clin. North. Am.* 17(4), 605–612 (1986)
80. Skalski, K., Kwiatkowski, K., Domanski, J., Sowinski, T.: Computer-aided reconstruction of hip joint in revision arthroplasty. *Journal of Orthopaedics and Traumatology* 7(2), 72–79 (2006)
81. Crawford, H.V., Unwin, P.S., Walker, P.S.: The CAD/CAM contribution to customized orthopaedic implants. *Proc. Inst. Mech. Eng. [H]* 206(1), 43–46 (1992)
82. Dunne, N.J., Orr, J.F.: Development of a computer model to predict pressure generation around hip replacement stems. *Proc. Inst. Mech. Eng. [H]* 214(6), 645–658 (2000)

83. Glinkowski, W., Wojnarowski, J.: Effect of calcar femorale upon the strength of proximal end of the femur: modeling with Finite Element Method. *Med. Sci. Monit.* 4(suppl. 2), 114–115 (1998)
84. Glinkowski, W., Ciszek, B.: Anatomy of the Proximal Femur -geometry and architecture. Morphologic investigation and literature review *Ortopedia, traumatologia, rehabilitacja* 4(2), 200–208 (2002)
85. Glinkowski, W., Wojnarowski, J.: Finite element modeling of strength of proximal femoral end during osteoporosis. *Postępy Osteoartrologii* 7, 61–66 (1995)
86. Glinkowski, W.: *Advances in International Telemedicine and eHealth* (Editor), Medipage, Warsaw, vol. 1 (2006)
87. Baruffaldi, F., Maderna, R., Ricchiuto, I., Paltrinieri, A.: Orthopaedic specialists' acceptance of videoconferencing consultations. *J. Telemed. Telecare.* 10(1), 59–60 (2004)
88. Vladzomyrsky, A.V.: Our experience with telemedicine in traumatology and orthopedics. *Ulus. Travma. Acil. Cerrahi. Derg.* 0(3), 189–191 (2004)
89. Baruffaldi, F., Gualdrini, G., Toni, A.: Comparison of asynchronous and realtime teleconsulting for orthopaedic second opinions. *J. Telemed Telecare* 8(5), 297–301 (2002)
90. Ricci, W.M., Borrelli, J.: Teleradiology in orthopaedic surgery: impact on clinical decision making for acute fracture management. *J. Orthop. Trauma.* 16(1), 1–6 (2002)
91. Tachakra, S., Hollingdale, J., Uche, C.U.: Evaluation of telemedical orthopaedic specialty support to a minor accident and treatment service. *J. Telemed. Telecare.* 7(1), 27–31 (2001)
92. Ricci, W.M., Borrelli, J.: Teleradiology in orthopaedics. *Clin. Orthop. Relat. Res.* (421), 64–69 (2004)

Computer-Aided Diagnosis: From Image Understanding to Integrated Assistance

Artur Przelaskowski

Institute of Radioelectronics, Warsaw University of Technology Nowowiejska 15/19,
Warszawa, Poland
arturp@ire.pw.edu.pl

Summary. This paper presents a status of the computer-aided diagnosis (CAD) in a context of nowadays challenges and limitations of advanced digital technologies in radiology, medical imaging systems and networked medical care. Computer-assisted interpretation of radiological examinations requires flexible more/less formal image modeling, reliable numerical descriptors of diagnostic content, indicators of image accuracy, and first of all effective methods of image understanding. Moreover, it is important to base the computer assistance design on understanding of human determinants of diagnosis, characteristics and enhancements of observer performance and dynamic platform of medical knowledge – a formal description of semantic image content, i.e. ontology. To make it fully useful, computer-based aid tools are integrated into networked radiology environment based on PACS/RIS/HIS/teleradiology systems interfaced to diagnostic workstations. The general concepts of CADs were exemplified inter alia with breast cancer and brain stroke diagnosis applications.

1 Introduction

Accurate image-based diagnosis depends on the quality of both the image acquisition and the image interpretation. While the development of medical imaging systems over the past century has been truly revolutionary because of tremendous advances in image detector systems and computer technology, methods of interpretation have only recently begun to benefit from advances in computer technology. Medical imaging breaking technologies has enormous potential to contribute to the improvement of health care and medicine in a near future. Imaging has come to include not only diagnostic methods but also treatments using image-guided methods. Increasingly, it depends not only upon the primary diagnostic technologies based on typical image acquisition, reconstruction, visualization and analysis, but also on "surrounding" technologies including information science and computational intelligence, networking, image archiving, storage and communication, contrast agent development, instrumentation, and treatment using physical energies etc. [1].

In consequences of rapid development, radiologists and clinicians routinely employ a variety of imaging modalities in daily diagnostic practice on a huge scale. Therefore, the fundamental and vital aspect is the correct and reasonably

fast interpretation of collected image data, referring to the physician's knowledge of typical, healthy and pathological anatomy and physiology of examined organs and structures, completed by experience and cognitive intuition. Full process of radiological interpretation, i.e. the understanding and assessment of medical image content, involves image-based detection of disease, defining disease extent, determining etiology of the disease process, assisting in designing of the clinical management plans for the patient, based on imaging findings, and following response to the therapy. Unfortunately, the interpretation of image examinations is still almost exclusively the work of humans what is expected to be changed in the next decades because of computer-based support.

Because the initial and key constituents of diagnosis are the true detection and defining of the disease, full understanding and accurate assessment of image content including also physical and technological conditioning is a key issue of successful exploitation of imaging capabilities. Diagnostic content is only a part of a given image information that is more complicated, ambiguous, technology-dependent, related to external, complementary data sources. Furthermore, information overload caused by increased volume of diversified medical image information, almost unlimited number of imaged object variations and known limitations of the human observers because of subjective in nature human factors cause the lack of stable standards in diagnosis. Thus, solid experience and cognitive intuition still play crucial role in the diagnosis.

2 Rationales

2.1 Human Limitations

Medical image perception is a very complex activity involving interplay between vision and cognition, and requiring detection, classification and actionable decision tasks performed on highly variable objects (i.e. lesions). But humans are limited in their ability to detect and diagnose disease during image interpretation due to their non-systematic search patterns and to the presence of structure noise masking normal and abnormal anatomical background. Observer variation is due to many factors – the degree of individual training, experience, and interest, imaging equipment and the quality of technical work performed. The remark that an experienced physician usually sees a visible lesion clearly but there are times when he does not, is still up-to-date. This is a baffling problem, apparently partly visual and partly psychological. It constitutes the still unexplained human equation in diagnostic procedures.

Observer performance improvement, i.e. increased stability, perfection, repeatability and reliability, is the main purpose of computer assistance. It starts with clarification (standardization, normalization) in more objective description of normal and abnormal findings including variability of symptoms regardless of acquisition parameters and technical conditions. Better understanding of content visualization, impact of the environment and reader fatigue, and better understanding and control of image quality issues including the physics, psychophysics and diagnostic

(semantic) measures lead to "intelligent", assistant interfaces, clearer rules of diagnosis, and generally improved methodology of diagnosis.

2.2 Computer-Aided Diagnosis

Objective support of computer-based aiding tools that storage, retrieve, communicate, select, emphasize, analyze, recognize, understand and visualize image-based diagnostic content can make significant improvements in the process of diagnosis and treatment of illness. Computer-based aid is sought to be an essential remedy for challenges of revolutionary changes in the medical imaging world. Supported radiological interpretation is expected to be more reliable and objective, repetitive, time- and even cost effective.

Generally, computer-aided diagnosis (CAD) is defined as a diagnosis that is made by a radiologist who uses the output from a computerised understanding of medical images as a assistance or "second opinion" in detecting lesions and in making diagnostic decisions. The final diagnosis is made by the radiologist [2]. Effective methods of CAD should be based on the fundamental content modeling concepts taking into consideration the following key issues:

- computational image descriptors:
 - flexible, reliable and specific numerical description of diagnostic image content: meaning of local regions, detailed structures, textures and global image features,
 - examination accuracy measures, diagnostic quality estimates, quality control tools,
- formalized medical knowledge:
 - characteristics of human possibilities and limitations due to image perception and understanding,
 - reliable observer performance characteristics – a methodology of interpretation: rules and protocols of effective diagnosis,
 - formalized medical knowledge platform: objective taxonomies and complete ontologies,
- semantic image understanding:
 - methods of semantic technologies,
 - content-based image indexing and retrieval (CBIR),
 - adjusted methods of image processing, analysis and synthesis, pattern recognition and understanding,
 - hierarchical and flexible cognitive resonance based on computational image descriptors and formalized medical knowledge,
- integration of medical information systems with advanced interfaces of diagnostic workstations and aiding tools
 - PACS/RIS/HIS/tele integration,
 - CAD/CBIR/ontology integration,
 - integrated user interfaces of every accessible information sources, systems and tools,
 - diagnostic workstation integration.

3 CAD Statement and Development

The following important issues were reported to underline CAD specificity – from image understanding to integrated assistance.

3.1 Computational Image Descriptors

Semantic image understanding concept means additional perspectives on the same image, taking into account the meaning and significance of its content. It is originated in the conceptions of semantic information theory¹ and is the essence of diagnostic image information.

Generally, semantic understanding means additional perspectives on the same image taking into account the meaning, and significance of its content. It is originated in the conceptions of semantic information theory and is the essence of diagnostic image information. Semantic content descriptors, extractors and empowers are fundamental for automatic interpretation of diagnostic image information. High-level analysis and recognition includes the use of computationally intelligent techniques, functional analysis of data, approximation theory methods and human visual models for image interpretation and human-following understanding. The relationship between image components, objects and patterns, its context – that play an important role in diagnosis – is related to and completed with a priori knowledge gained from a range of sources. Computational image interpretation are expected to be additional, semantic eyes of diagnosis.

Nowadays, semantic technologies have the power to revolutionize the IT world. In a heterogeneous world of ubiquitous information flow they allow a flexible and seamless integration of applications and data sources (i.e. radiological chain of patient, imaging system, interpretation, and therapy). They provide an intelligent access, understand context and content, give answers and generate knowledge including as objective as possible object description (e.g. lesion, structure relation, pathology features). Semantic annotation of images is a key concern for the newly emerged applications of semantic multimedia. Machine processable descriptions of images make it possible to automate a variety of tasks from search and discovery to composition and collage of image data bases. Automatic understanding to improve machine interpretation of the images is more demanding challenge. Image accuracy estimates, numerical quality measures and semantic information metrics reflect automatic understanding of the image content following subjective, semantic interpretation of the information. Numerical descriptors of diagnostic content quality are searched, designed, verified and used according to medical expert requirements. The expected results are computational models of image semantics.

3.2 Semantic Visualization

Research in human perception of lesion symptoms requires objective methodologies for optimal image presentation, i.e. image diagnostic quality enhancement,

¹ <http://plato.stanford.edu/entries/information-semantic/>

semantic information selection, description and extraction, fitting of display parameters. Psycho-physical models for detection of abnormalities based on the understanding of what is desired by an image observer, what properties of radiological images are the most useful in their interpretation, and how these properties can be enhanced to improve the accuracy of interpretation, are sought. Optimised visualization communicates selected, extracted and empowered diagnostic image features (i.e. lesion symptoms) to the human visual system instead of large amounts of disordered information.

Exemplary application of improved semantic image perception is MammoViewer² developed in our lab. Wavelet-based multiscale decomposition was used for effective mammogram data denoising and enhancement to extract suggestively microcalcifications and masses. Another example is Stroke Display³ implied as a kind of intelligent data visualization method that communicates selected, extracted, and enhanced ischemic hypodensity signs to the observers, especially for "radiologically silent" cases (really difficult to diagnose). It complements conventional CT display with additional display highly specific in infarct cases. As computer-assisted interpretation tool Stroke Display was designed to uncover, model and enhance signatures of ischemia. Multiscale transformations was used to analyze image content basing on spatially distributed soft tissue properties over different scales and subbands. Pathology signs may be effectively strengthened, extracted and identified through hierarchical local data processing.

3.3 Medical Knowledge Platform

Primary goal of ontology is to represent effectively a domain knowledge, adequately and exhaustively define relevant concepts, object characteristics and relationships between them, to provide a common, standardized vocabulary comprehensible by humans and machines by which users and computer systems can communicate. Thus, ontology means systematization, objectification and verification, knowledge base of the model populated with concept's instances constitutes standard diagnostic knowledge database. Ontologies are the foundation of the Semantic Web, where integration and interoperability of heterogeneous sources of information is needed [3]. Ontologies also form the basis foundation of evidence-based-medicine and standardization efforts [4].

As an example, mammographic ontology constructed by [5] has three main goals: – to provide standard vocabulary and formal, exhaustive definitions of concepts for description and interpretation of mammograms; – to model mammography report; – to use ontology as specification for designing the database of mammography reports and graphical editor for pathologies description. Other designed applications of the mammographic ontology are: educational tasks, as an assistant tool for diagnosis in mammography, and content-based indexing of mammograms database. Knowledge of ontology construction has been extracted from three sources: corpus of routine, free-text mammography reports, long-term

² <http://www.ire.pw.edu.pl/MammoViewer/>

³ <http://aidmed.pl>

interviews and consultations with radiologists at local hospital and careful wide range analysis of medical literature.

3.4 Cognitive Resonance for Image Understanding

The important aspect of computer-based aid development and technological progress is the understanding of the relevant semantic contents of the radiological examinations on the basis of numerical features extracted from the image data [6]. Occurring problem of the semantic gap between the low level numerical descriptors and the high level interpretation of an image based on medical knowledge poses new challenges and needs. Because low level descriptors cannot be uniquely associated with any other meaningful label unless explicitly declared or derived as the outcome of a classification procedure, retrieval or automatic recognition and diagnosis based on knowledge level constructs is a non-trivial task to achieve in general. Suggested solution is cognitive resonance of more complex and packed descriptors with objectified, simplified and formalized knowledge bases.

The subject of interest is the application of the cognitive-based approach for intelligent semantic analysis, allowing the automatic description of important diagnostic features of analyzed images. The most important part of this analysis depends on the "cognitive resonance" process, in which the features of real images are compared with some kind of expectation taken from the knowledge base, containing the characteristics of the pathological cases originated from medical practice. The importance of cognitive resonance in medical image understanding was noticed and confirmed by the research of R.Tadeusiewicz and M.Ogiela [7].

Some important approaches were based on a special kind of image description language and grammar formalism. During the linguistic analysis of medical patterns, we can solve the problem of generalization of features of a selected image and obtaining semantic content description of the image [8].

From other perspective, the cognitive resonance may be approximated from the following issues:

- content low-level measurement as a signal extraction from case dependent complex masking conditions; low-level descriptors effectively characterized significant signal features are sought;
- content high-level clarifying as a knowledge systematization, objectified, standardized and hierarchized, i.e. accurate ontologies for selected examinations and modalities;
- content identification as the reliable integration of computed low-level semantics of the concrete examination and systematized semantics based on selected medical knowledge.

Consequently, computational understanding of medical image semantics is suggested to be based on the following items:

- extraction and numerical description of image diagnostic information optimized according to reliable semantic criteria through:
 - estimating local and global features correlated to diagnostic content,

- design of the descriptors and diagnostic quality/accuracy indicators, e.g. wavelet-based and other multi-scale sensing technologies may be used for representation, separation and recognition of the signal useful-in-diagnosis [9];
- experimental verification of diagnostic information measures with design of image subjective rating, pathology detection tests, ROC-based analysis etc.
- constituting objective, as formal as possible medical knowledge platform which is gold standard reference of meaning and significance; taxonomy and ontology-based technologies were applied to create formal, hierarchized information structure to extract:
 - inter-levels, inter-classes, and inter arguments dependences of the knowledge;
 - standardized, pathology descriptions, classified, complete feature characteristics of abnormalities, benign/malignant classifiers of the tumors etc.
- providing an access to the reference database, i.e. realization of dynamic, easy-to-access source of up-to-date diagnostic information based on knowledge-by-example concepts through:
 - effective, easy-to-use networking technologies (e.g. grids), image archives and distribution protocols, retrieval engines, integrated interfaces;
 - content-based image indexing and retrieval, fast and effective dictionary structures, reliable similarity measures of image content;
- integration and interfacing of multiple stages of the iterative cognitive resonance process which means optimization of recognition, classification and the similarity identification according to the semantic criteria; the essence is modeling of numerical image description to be fitted as reliably as possible to the standardized, objectivated, formalized and up-to-date diagnostic knowledge.

Advanced techniques of computational intelligence were applied in order to recognize and understand semantic content of the images. Computational intelligence (NN, fuzzy sets used as classifiers and decision support) can improve the intellectual behavior of machines by better description, extraction and cognitive-based identification of semantic information. Making semantic data interpretation more amenable to computational methods is the main subject of considered research. Formal description of abnormalities, interpretation criteria and protocols, the relationships (properties) between these concepts resonated to computational methods of determining of diagnostic accuracy indicators (recognition and understanding in diagnostically important terms), as selective, representative and compact as possible, are well readable by both human beings (radiologists) and machine agents (CAD systems).

3.5 Content-Based Indexing of Medical Images

Content-based retrieval will likely become more commonly used for medical information retrieval. CBIR has been one of the most intense research areas in the

field of computer vision over the last years. The availability of large and continuously growing amounts of visual and multimedia data, and the development of the Internet point the need to create content-based access methods that offer more than simple text-based queries.

It needs to be stated that the purely visual image queries as they are executed in the computer vision domain will not most likely be able to ever replace text-based methods as there will always be queries for all images of a certain patient, but they have the potential to be a very good complement to text-based search based on their characteristics. Still, the problems and advantages of the technology have to be stressed to obtain acceptance and the use of visual and text-based access methods up to their full potential. A scenario for hybrid, textual and visual queries was proposed in the CBIR2 systems [10].

The benefits from CBIR engine for PACS user seem to be incontrovertible. Even if CBIR is relatively simple and able to distinct only modalities and some body parts – it might be a good supplement to classic text query engine, giving an opportunity to verify misclassified DICOM-based information, which happens in clinical practice. Nowadays, the CBIR improvement is based on semantic understanding of technologies in the context of semantic-based knowledge management as promising development factor of IT world. In a heterogeneous reality of ubiquitous information flow, they allow a flexible and seamless integration of applications and data sources (i.e. radiological chain of patient, imaging system, diagnostic interpretation and therapy). They provide an intelligent access, understand context and content, give answers and generate knowledge including as objective as possible object description (e.g. lesion, structure relation, pathology features).

The inclusion of visual and semantic features into medical studies is an interesting point for several medical research domains. Content-based features do not only allow the retrieval of cases with patients having similar diagnoses, but also cases with visual and semantic similarity but with different diagnoses. In teaching, it can help lecturers as well as students to browse educational image repositories and visually inspect the results found. Image interpretation is expected to be improved through the use of content-based access methods to the existing large repositories of diagnoses cases. Reference database of image examinations indexed by content is an exciting challenge for semantic analysis of image data: it provides the examples of clinically verified cases of pathologies. The retrieval of diagnostically similar exams over as wide as possible distributed databases of reliably labeled cases was used to support diagnosis of cases difficult to assess. Image retrieval techniques are based on similarity matching of diagnostic image content. Between interesting algorithms, those based on multiresolution histogram matching, multiscale data distributions and local structure correlations, subband texture analysis are worth mentioning.

Our exemplar CBIR system [11] was based on mammographic ontology, semantic descriptors of abnormalities and CAD support. The precision of image retrieval depends on CAD sensitivity, lesion characteristics, precision of breast tissue differentiation, the estimates of local and global features of the image,

and textual and numerical descriptors based on ontology (e.g. texture, shape, outline, localization).

4 Integrated Interpretation Assistance

Radiology is extremely susceptible to computer-based integrated support in decision-making. Integration of databases, decision-aiding tools, systems, networks can help to determine what information is needed that a user does not have. Information such as the reason for the imaging examination being ordered by the referring clinician, laboratory data, and patient clinical history, in addition to cross-specialty and cross-modality imaging should be integrated and easily accessed. Decision support tools will provide more information to end-users, but need to be more fully integrated into the PACS database.

The purpose of designed computer-assisted medical systems is the completeness of supplied aid to improve the diagnosis. However, visualization and navigation of all accessible medical information requires efficient selection of all the information necessary to make effective clinical decisions without distracting the user, followed by information synthesis. Automation by integration can improve database information quality, as well as facilitate improved user interfaces, computer-aided tools, and preemptive detection of errors before they propagate. The importance of provided information quality, CAD-based support and tele-radiology, flexible adjustment to various needs, and easily accessible examples of pathology cases, quick and precise retrieval, are crucial.

End-users want more functionality but simpler user interfaces. Intelligent user interfaces (IUI) are designed with the following features and demands in mind: efficient and reliable visualization of all useful information, adoption to human performance and limitations, intuitiveness and consistency, flexibility and easy configurability (to accommodate different types of users and different types of imaging examinations), appeal and idiot-proof. IUI is based on effective models and metrics for function task lists to read different case types, workflow guidelines, common language and best practice for hanging protocols.

The integration is essential to improved diagnostic data evaluation, decision making and finally high-quality patient care. However, data gathering, flow and exchange standardization, IUI design and communication of the results is a great challenge for life science community. An integrated tele-information system for clinical and research purposes developed in collaboration between Warsaw University of Technology, Wolski Hospital and software houses⁴ is an example of integrated assistance.

5 Conclusions

Generally, computer-assisted interpretation of image exams is a complex problem of physician performance support, integrating different methods and forms of data processing, visualization, analysis, classification, understanding and navigation.

⁴ <http://telemedycyna.evernet.com.pl/system/>

Semantic understanding of medical images can be improved by interdisciplinary integration of description methods, intelligent numerical models, and implemented tools but first of all by the increased unification of the meaning and significance of assessed pathology symptoms, computational understanding of abnormality indications and technical conditioning of diagnosis.

Generally, computer-aided diagnosis requires integration of the computational methods, information systems and objective medical knowledge systematization based on the following issues:

- **reliable determining of semantic contexts**, extraction and numerical description of image diagnostic content by semantic local and global feature selection, analysis, compaction with redundancy reduction and understanding to formulate automatic interpretation;
- **constituting objective, as formal as possible, medical knowledge platform** which is a gold standard reference of the meaning and significance for diagnosis;
- **adaptive cognitive recognition of medical knowledge (patterns of abnormalities)**, cognitive resonance process to synchronize computational efforts with formalized medical knowledge based on recognition, classification and the similarity identification according to the semantic criteria; the essence is the modeling of low-level numerical descriptors resonated to normalized, hierarchized, objectified diagnostic rationales,
- **development of integrated information systems** with reference knowledge, world-wide access based on distributed databases, content-based image indexing, retrieval and communication.

CAD systems can be improved by more effective mathematical modeling of semantic content resonated to formal domain knowledge basing on integrated assistance that support necessary information access, ordering, reference and management. Even though a design of human-machine interface is the most challenging aspect of computer-aided interpretation of medical image exams, both the technology improvements and increased objectivity of human performance criteria and procedures are the fundamental conditions for expected success.

References

1. Future needs for medical imaging in HC: Medical Imaging Technology Roadmap (2000), http://www.ic.gc.ca/epic/site/mitr-crtim.nsf/en/h_hm00038e.html
2. Giger, M.L.: Computer-aided diagnosis in medical imaging - A new era in image interpretation. World Markets Research Centre, Tech. Rep. (2000)
3. Maedche, A.: Ontology learning for the semantic Web. Kluwer Academic Publishers, Dordrecht (2003)
4. Pisanelli, D.M., et al.: An ontological approach to evidence-based medicine and meta-analysis. In: Proc. Medical Informatics Europe 2003, IOS Press, Amsterdam (2003)
5. Podsiadly-Marczykowska, T., Guzik, A.: Mammography ontology, model structure, definitions and conception instances. *Bio-Algorithms and Med.* 1(1), 247–252 (2005)

6. Tadeusiewicz, R.: Automatic understanding of signals. In: Proc. of the IIPWM 2004 Conference on Intelligent Information Processing and Web Mining, Springer, Berlin–Heidelberg–New York (2004)
7. Tadeusiewicz, R., Ogiela, M.: Medical image understanding technology: artificial intelligence and soft-computing for image understanding. *Studies in Fuzziness and Soft Computing*, vol. VIII, p. 156. Springer, Berlin-Heidelberg-New York (2004)
8. Ogiela, M., Tadeusiewicz, R.: Nonlinear processing and semantic content analysis in medical imaging - a cognitive approach. *IEEE Trans. Instrum. Meas.* 54(6), 2149–2155 (2005)
9. Chan, A.K., Peng, C.: *Wavelet for sensing technologies*. Artech House, Inc. (2003)
10. Muller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. *Int. J. Med. Inform.* 73, 1–23 (2004)
11. Boninski, P.: *Medical image indexing for digital radiology*. PhD thesis, Warsaw University of Technology (2007)

Biomedical Structures Representation by Morphological Spectra

Juliusz L. Kulikowski, Małgorzata Przytułska, and Diana Wierzbicka

Institute of Biocybernetics and Biomedical Engineering Polish Academy of Sciences,
Ks. Trojdena Str. 4 02-109 Warsaw, Poland
{jlkulik, gosia, diana}@ibib.waw.pl

Summary. It is considered the problem of representation of irregular structures, typical in biomedical images, by morphological spectra. Some basic properties of morphological spectra have been reminded. Then several possible methods of morphological spectra presentation are illustrated by a numerical example. The problem of representation of selected classes of irregular structures is considered and illustrated by the examples of vertically elongated, compact and branching structures representation. It is shown that for each class of irregular structures a hierarchy of spectral components indicating their role in representation of the class can be statistically established.

1 Introduction

Recognition and localization of typical biomedical structures is a basic step to computer-assisted image analysis both in biological investigations and in medical diagnosis. Shapes and textures of biomedical objects in formal sense are not strongly defined. They rather form some classes of structures that only roughly can be described in the terms of geometry [1], Fourier spectra [2], Markov processes [3], formal linguistics [4], relations theory [5, 6], etc. In each of the above-mentioned cases individual objects (e.g. neural cells, leukocytes, cardiac arteries, lung tumors, cerebral ischaemic areas, etc.), are different and their general description as classes of formal objects is only approximately possible. We call *regular* the shapes whose class can be completely and exactly defined in the terms of a certain mathematical theory. In this sense regular is a class of all circles of limited diameters, a class of all real continuous periodic functions of a fixed period, etc. Unlike the regular ones, the classes of biomedical structures cannot be by any mathematical theory determined without leaving some their instances outside and, at the same time, without admitting some false structures to belong to them. Essential properties of biomedical structures are connected with their biological origins and living functions and as such by formal models they can be only approximately described. The models can be assessed from two basic points of view: 1st of accuracy, and 2nd of complexity of biomedical structures description. In this paper utility of morphological spectra as tools for visualized biomedical structures description and recognition is presented. Basic notions and properties of 2D morphological spectra have been published in [7, 8]. They also are shortly reminded in Sec. 2 of this paper. In Sec. 3 examples

of morphological spectra of selected 2D structures are presented. In Sec. 4 the problem of concise representation of classes of biomedical structures in the terms of morphological spectra is considered. Final conclusions are presented in Sec. 5 of the paper.

2 Morphological Spectra

There are considered 2D monochromatic images represented by bit-maps of $2^m \times 2^n$ size, m and n being some fixed natural numbers. Morphological spectra of images represented by the bit-maps form a hierarchical, multi-level structure; the number $k^* = \min(m, n)$ determines the highest level of the structure. The k -th level spectral components, $0 \leq k \leq k^*$, are calculated on $2^k \times 2^k$ square image segments called basic windows. For this purpose the image is partitioned into $2^{m-k} \times 2^{n-k}$ adjacent basic windows and the values of a given spectral component for each basic window are independently calculated. The k -th level morphological spectrum consists of 2^{2k} components. Any spectral component of a total image takes the form of a $2^{m-k} \times 2^{n-k}$ real matrix. The components are denoted by k -element strings of symbols Σ , V , H , X and, respectively, can be lexicographically ordered. The 0^{th} level spectrum is by definition identical to the original image. The 1^{st} level morphological spectrum consists of the components Σ , V , H and X only. The 2^{nd} level morphological spectrum consists of 16 components: $\Sigma\Sigma$, ΣV , ΣH , ΣX , $V\Sigma$, VV , VH , VX , $H\Sigma$, HV , HH , HX , $X\Sigma$, XV , XH and XX , etc. The components of morphological spectra can be represented by a tree whose nodes are assigned to the components and edges indicate the relationships of direct hierarchical submission of lower level to the higher level spectral components. The root of the tree is assigned to the original image. An upper part of the tree is shown in Fig. 1.

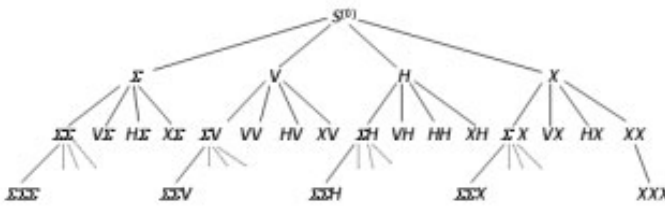


Fig. 1. Hierarchical tree of morphological spectra

If we put

$$\mathbf{U}^{(k)} = [\xi_{1,1}, \xi_{1,2}, \dots, \xi_{1,N}, \xi_{2,1}, \dots, \xi_{2,N}, \dots, \xi_{N,1}, \dots, \xi_{N,N}] \quad (1)$$

where $N = 2^k$, denotes pixel values of a basic window arranged in a linear (vector) form then its morphological spectrum can also be represented in vector form $\mathbf{V}^{(k)}$ and calculated as:

$$(\mathbf{V}^{(k)})^{tr} = \mathbf{M}^{(k)} \cdot (\mathbf{U}^{(k)})^{tr} \tag{2}$$

where tr is a symbol of vector (matrix) transposition, $\mathbf{M}^{(k)}$ is a square $2^{2k} \times 2^{2k}$ spectral matrix. For $k = 1$ it takes the form:

$$\mathbf{M}^{(1)} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 \end{bmatrix} \tag{3}$$

It can easily be proven that $\mathbf{M}^{(1)}$ has the following property of orthogonality:

$$\mathbf{M}^{(1)} \cdot (\mathbf{M}^{(1)})^{tr} = 2^2 \cdot \mathbf{I}^{(4)} \tag{4}$$

where $\mathbf{I}^{(4)}$ is a 4-th order unity matrix. This means that:

$$(\mathbf{M}^{(1)})^{-1} = 2^{-2} \cdot (\mathbf{M}^{(1)})^{tr} \tag{5}$$

It follows from (2) and (5) that:

$$\mathbf{U}^{(1)} = 2^{-2} \cdot \mathbf{S}^{(1)} \cdot \mathbf{M}^{(1)} \tag{6}$$

The property (4) can easily be extended on $k > 1$ and, as a consequence, formula (6) takes a more general form:

$$\mathbf{U}^{(k)} = 2^{-2k} \cdot \mathbf{S}^{(k)} \cdot \mathbf{M}^{(k)} \tag{7}$$

where $\mathbf{S}^{(k)}$ is the k -th level morphological spectrum of $\mathbf{U}^{(k)}$. The principles of $\mathbf{M}^{(k)}$ construction on the basis of k -th level morphological spectral masks are described in [8].

Below the spectral matrix $\mathbf{M}^{(2)}$ is shown:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$\Sigma\Sigma$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ΣV	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1
ΣH	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1	1	1
ΣX	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1
$V\Sigma$	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1
VV	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1
VH	1	1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1
VX	1	-1	-1	1	-1	1	1	-1	1	1	-1	1	-1	1	1	-1
$H\Sigma$	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1
HV	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1
HH	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1
HX	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	1	-1	1	-1
$X\Sigma$	-1	-1	1	1	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1
XV	1	-1	-1	1	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1
XH	1	1	-1	-1	-1	-1	1	1	-1	-1	1	1	1	1	-1	-1
XX	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1

It can be used to calculation of spectral components on 4×4 basic windows. Higher-level spectral matrices are rather inconvenient to be graphically presented: $\mathbf{M}^{(3)}$ is a 64×64 matrix, $\mathbf{M}^{(4)}$ a 256×256 one, etc. However, due to simple principles of their creation their components can be easily calculated.

3 Spectral Representation of Single Irregular Objects

It has been mentioned above that irregular objects given in the form of $2^m \times 2^n$ bit-maps can be represented by k -th level morphological spectra for $k = 1, 2, \dots, k^*$. Each such representation contains full information about the original image and the original image due to the formula (7) from any k -th level morphological spectrum can be restored. However, equivalence of spectral representations does not mean that in a given situation all representations are equally useful. As an example let it be considered a binary image of 4×4 pixels size, of an irregular object shown in Fig.2 together with its bit-map.

$$a) \begin{array}{cccc} \circ & \circ & \bullet & \bullet \\ \circ & \bullet & \circ & \bullet \\ \circ & \circ & \circ & \bullet \\ \bullet & \bullet & \circ & \circ \end{array} \quad b) \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Fig. 2. Example of a) a binary image of an irregular object and b) of its bit-map

For this image the 1^{st} and 2^{nd} level morphological spectra only can be calculated. They take the following forms:

a / the 1^{st} level spectrum in components separated form:

$$\Sigma = \begin{bmatrix} 3 & 1 \\ 2 & 3 \end{bmatrix}, \quad V = \begin{bmatrix} -1 & -1 \\ 0 & -1 \end{bmatrix}, \quad H = \begin{bmatrix} -1 & 1 \\ -2 & 1 \end{bmatrix}, \quad X = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix};$$

b/ the 1^{st} level spectrum in compact form:

$$S^{(1)} = \left[\begin{bmatrix} 3 & -1 & -1 & 1 \\ 2 & 0 & -2 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 & 1 \\ 3 & -1 & 1 & -1 \end{bmatrix} \right];$$

c/ the 2^{nd} level spectrum in compact form:

$$S^{(2)} = [9, -3, -1, 1, 1, -1, 5, -1, 1, -3, -1, -3, -3, 1, -1, 1],$$

the components of $S^{(2)}$ being ordered as mentioned above.

The last presentation form is the most concise one and convenient to a description of larger classes of irregular objects.

4 Spectral Representation of Classes of Objects

Spectral representation of classes of irregular objects can be given explicitly in the form of lists of spectra of the objects. However, the role of spectral components in description of the objects is different and depends on the shapes of the objects. In order to investigate this problem there were considered the classes of vertically elongated, compact and branching irregular objects. For each class 24 instances of irregular objects were generated; some examples of objects of the classes being shown in Fig. 3.

For the objects their 2^{nd} level morphological spectrum were calculated. Then the spectra were examined from the point of view of weights of spectral components occurring in the spectra of a given class. For this purpose absolute values of each component were normalized with respect to the value of the component $\Sigma\Sigma$ and summed over all instances of objects of the given class. The results were averaged over the set of instances. The results are given in Table 1 together with standard deviations of the estimated mean values.

For better visualization of the results they are presented in a graphical form in Figs. 4, 5 and 6.

It can be observed that the relative role of spectral components in the objects of a given class representation is different. In the case of vertically elongated structures the differences are particularly high and it is possible to indicate three groups of spectral components playing a crucial role in this class of irregular objects characterization: 1^{st} , the most important group consists of the $V\Sigma$ and

Table 1. Averaged normalized weights of spectral components

Class of objects	Averaged normalized weights of spectral components							
	$\Sigma\Sigma$	ΣV	ΣH	ΣX	$V\Sigma$	VV	VH	VX
Vertical elongated	1.000 ± 0.0	0.486 ± 0.365	0.053 ± 0.146	0.320 ± 0.264	0.889 ± 0.267	0.597 ± 0.364	0.111 ± 0.133	0.320 ± 0.242
Compact	1.000 ± 0.0	0.183 ± 0.096	0.142 ± 0.093	0.186 ± 0.130	0.415 ± 0.249	0.504 ± 0.136	0.254 ± 0.171	0.271 ± 0.184
Branching	1.000 ± 0.0	0.216 ± 0.410	0.185 ± 0.223	0.272 ± 0.238	0.240 ± 0.156	0.229 ± 0.245	0.230 ± 0.228	0.274 ± 0.212

Table 2. Averaged normalized weights of spectral components

Class of objects	Averaged normalized weights of spectral components							
	$H\Sigma$	HV	HH	HX	$X\Sigma$	XV	XH	XX
Vertical elongated	0.053 ± 0.045	0.430 ± 0.330	0.053 ± 0.146	0.389 ± 0.292	0.139 ± 0.254	0.403 ± 0.280	0.153 ± 0.131	0.305 ± 0.273
Compact	0.422 ± 0.073	0.250 ± 0.131	0.494 ± 0.195	0.211 ± 0.152	0.257 ± 0.251	0.318 ± 0.227	0.315 ± 0.201	0.318 ± 0.258
Branching	0.172 ± 0.126	0.321 ± 0.243	0.263 ± 0.172	0.270 ± 0.199	0.235 ± 0.230	0.216 ± 0.324	0.267 ± 0.244	0.229 ± 0.195

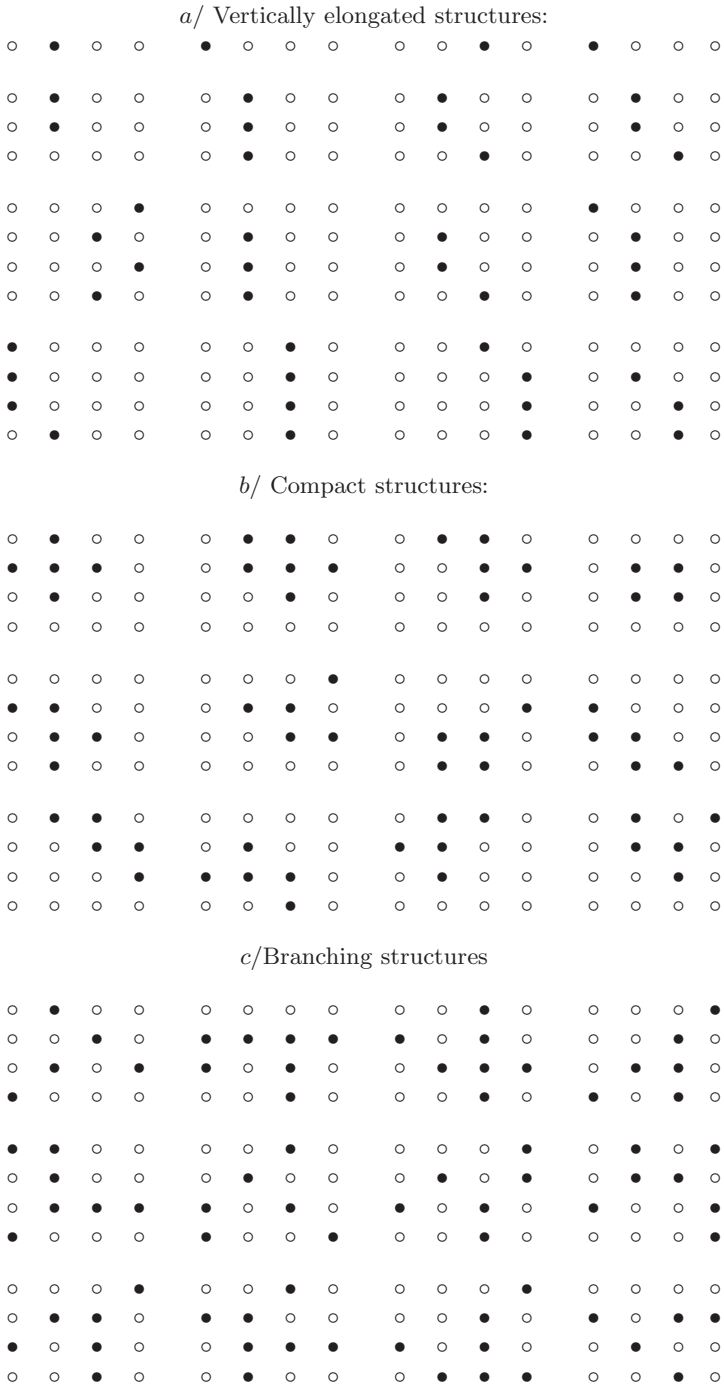


Fig. 3. Examples of irregular structures used in the experiment

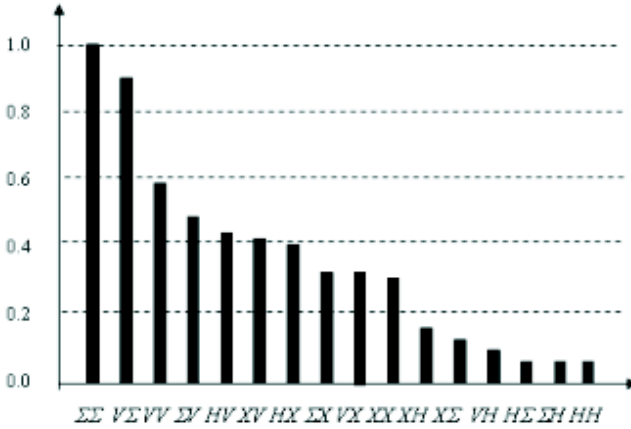


Fig. 4. Averaged normalized weights of morphological spectral components describing irregular elongated structures

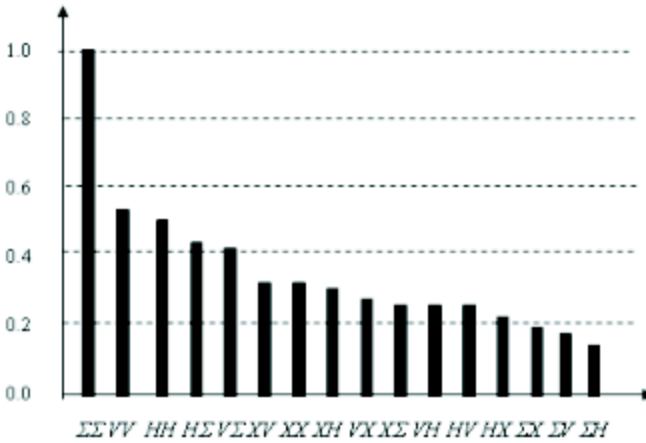


Fig. 5. Averaged normalized weights of morphological spectral components describing irregular compact structures

VV components, 2^{nd} , the middle-importance group consists of the HH , $H\Sigma$, $V\Sigma$, XV , XX and XH components, the rest belong to the group of components playing less important role in characterization of this class of irregular objects. It should be remarked that due to the property of symmetry of morphological spectra the above-given results can also be extended on a class of horizontally elongated irregular objects: in such case the tags of spectral components should be only changed by putting symbol H instead of V and vice versa.

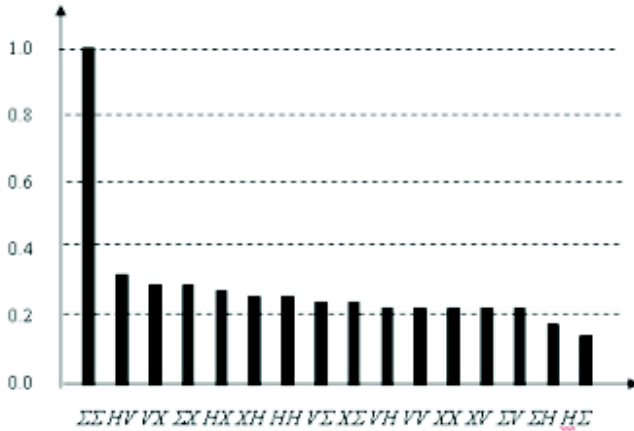


Fig. 6. Averaged normalized weights of morphological spectral components describing irregular branching structures

In two other classes of irregular objects a hierarchy of relative importance of spectral components can also be observed, however, the differences are not so large. In detection of compact irregular structures the relatively most important role play the components VV , HH , $H\Sigma$ and $V\Sigma$. In characterization of branching structures the relatively most important role play the components HV , VX , ΣX and HX , however, their dominating role in branching structures description is not so evident.

Let us remark that the above-presented experiment shows only that for certain classes of irregular objects certain spectral components play more important role than the other ones. However, the 2^{nd} level morphological spectra give rather limited possibilities to describe well some more sophisticated morphological structures because of limited size of their basic windows (4×4 pixels). Nevertheless, if the $V\Sigma$ and VV components play a crucial role in vertical structures description (similarly, $H\Sigma$ and HH in horizontal structures description) then this means that the higher-level components of the type $V\Sigma NN$ and $VV NN$ (respectively, $H\Sigma NN$ and $HH NN$), where NN denotes any string of symbols Σ , V , H , X , will also play a similar role in higher-level vertically elongated (respectively, horizontally elongated) structures characterization.

However, it also may happen that for certain classes of irregular structures no group of the most suitable structural components for the class characterization can be easily found. In our experiments the classes have been chosen in arbitrary way and the instances of the classes were generated manually, what means the classes of structures (like the real biomedical structures) had no strong mathematical sense. In such case examination of the role of separately taken spectral components may be not effective. Instead of this, co-occurrence of their pairs, triples, etc. in the spectra of examined structures should be taken into consideration. The problem needs more experimental investigations.

5 Conclusions

It has been shown that morphological spectra are a useful tool for irregular structures description. The role of spectral components in different in different classes of morphological structures characterization. The problem is of great importance both in morphological filters design, as well as in pattern recognition. However, the problem needs further and deeper investigation.

Acknowledgement. This work was supported by Ministry of Science and Higher Education of Poland, project nr. 4211/BT02/2007/33.

References

1. Zhu, Y.M., Gao, Y., Goutte, R., Amiel, M.: Textural Boundary Detection Using Local Spatial Frequency Analysis. In: Proc. 11th IAPR International Conference on Pattern Recognition, Hague, vol. III, pp. 53–56. IEEE Computer Society Press, Los Alamitos (1992)
2. Haddon, J.F., Boyce, J.F.: Texture Segmentation and Region Classification by Orthogonal Decomposition of Cooccurrence Matrices. In: Proc. 11th IAPR International Conference on Pattern Recognition, vol. I, pp. 692–695. IEEE Computer Society Press, Los Alamitos (1992)
3. Ojala, T., Pietikajnen, M.: Unsupervised Texture Segmentation Using Feature Distributions, Texture Analysis Using Pairwise Interaction Maps. In: Del Bimbo, A. (ed.) ICIAP 1997. LNCS, vol. 1310, pp. 311–318. Springer, Heidelberg (1997)
4. Loum, G., Lemoine, J., et al.: An Application of Wavelet Transform to Texture Analysis. In: Proc. of the 9th Scandinavian Conference on Image Analysis, Uppsala, vol. 1, pp. 583–590 (1995)
5. Xiaohan, Y., Yla-Jaaski, J.: Unsupervised Texture Segmentation Based On the Modified Markov Random Field Model. In: Proc. 11th IAPR International Conference on Pattern Recognition, Hague, vol. III, pp. 88–91. IEEE Computer Society Press, Los Alamitos (1992)
6. Smith, T.G., Lange, G.D.: Biological Cellular Morphometry-Fractal Dimensions, Lacunarity and Multifractals. In: Losa, G.A., Merlini, D., et al. (eds.) Fractals in Biology and Medicine, Birkhauser, Basel, vol. II, pp. 30–49 (1998)
7. Kulikowski, J.L., Wierzbicka, D.: A Method of Microvascular Systems Analysis Based on Statistical Texture Parameters Evaluation. *Biocybernetics and Biomedical Eng.* 23(3), 21–37 (2003)
8. Kulikowski, J.L., Przytulska, M., Wierzbicka, D.: Recognition of Textures Based on Analysis of Multilevel Morphological Spectra. In: IFMBE Proceedings of World Congress on Medical Physics and Biomedical Engineering, Seoul, pp. 2164–2167 (2006)

Medical Image Analysis Using Potential Active Contours

L. Pieta¹, A. Tomczyk¹, and P.S. Szczepaniak^{1,2}

¹ Institute of Computer Science, Technical University of Lodz,
Wolczanska 215, 90-924, Lodz, Poland

² Systems Research Institute, Polish Academy of Sciences Newelska 6,
01-447 Warsaw, Poland
tomczyk@ics.p.lodz.pl

Summary. Potential contours are methods for automatic image analysis. They can be interpreted as contextual classifiers that use expert knowledge and operate in supervised or unsupervised mode. In the present paper, potential contours adapted in the supervised way are examined on medical images.

1 Introduction

Medical imaging [1] is an important application area of image processing [2]. Within medical image processing, segmentation is one of the most difficult tasks. Originally, active contour methods were developed as tools for a low-level image segmentation with ability for use of high-level information [3, 4]. The main idea is to find an optimal contour in the space of considered contours representing certain region on the image. The search is performed in an evolution process (optimization) in which the given objective function, called energy, evaluates the quality of contour. As shown in [5, 6], contours are contextual classifiers of pixels (one part of pixels belongs to the interior and another one - to the exterior of given contour, and active contours are methods of optimal construction of classifiers. Potential active contour possesses ability to evolve with change of the location or number of control points, and with modification of parameters of potential functions. The search of optimal contour is performed by optimization of some performance index E called *energy* in the theory of active contours. In E almost any type of information can be used assuming that we are able to implement this information in the computer oriented form.

The search of the optimal contour may be driven in many ways - e.g. by use of simulated annealing or genetic algorithm which perform global search and do not use gradient. In our work we apply the first mentioned method.

Adaptation is another interesting and powerful mechanism [7]. Discrimination ability of a given contour is limited and it depends on the number of control points (assuming that other parameters are fixed). Flexibility of the potential active contours can be improved if we incorporate the change of the number

of control points into the optimization procedure. For example, we can start with small number of that points and add new ones, if necessary. The rate of misclassification in some area can be the reason for introducing a few new control points.

2 Preprocessing

External energy function in the [resented application operates on greyscale images. The value of contour energy depends on the intensity of each pixel. A contour with the lowest energy value is one that covers only black pixels of the image. A contour that covers only white pixels has the highest energy value. Therefore, the annealing mechanism, which uses external energy, adjusts the contour to dark pixels. For example, in Fig.1a the contour is adjusted to black pixels of the circle.

Preprocessing of the input picture, which consists in the blurring of black contours, improves segmentation. Every image undergoes a preparation process, during which the color of pixels is scaled so that their shades can change fluidly from white (the color of background) to black (the color of object). The algorithm

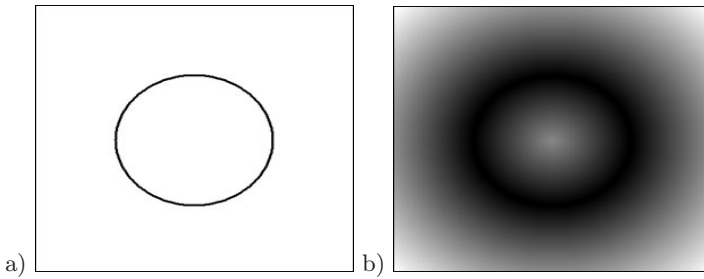


Fig. 1. (a) Test image of a circle consisting of black pixels. (b) Image of a circle during a preparation process (blurring of black contours).

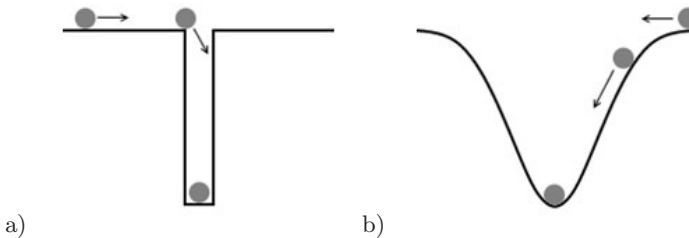


Fig. 2. (a) An image without blurring. The horizontal line represents high energy value. In search of low energy value, the contour should "fall into" a steep gap. (b) Blurred image, the energy value changes smoothly.

assigns each pixel a color in a greyscale, depending on the distance between the pixel and the nearest black point of an image, as illustrated by the formula:

$$P(x, y) = \frac{255\rho(x, y)}{\rho_{max}} \quad (1)$$

where $P(x, y)$ - the resultant intensity of light at a pixel specified by coordinates (x, y) , $\rho(x, y)$ - Euclidean distance between point (x, y) to the nearest black point of an object, ρ_{max} - the longest distance between the point of the image and the nearest black point of the object. The result is shown in Fig. 1b. The blurring of contours enhances the method's efficiency, because the contour behaves similarly to a sphere rolling down a slope into a valley, that is the lowest point - Fig. 2. Blurring of contours can be implemented in external energy function. However, energy value is calculated at every step of the algorithm, which means that an image is blurred with each iteration. During segmentation, the image does not change. Thus, it is more useful to blur the image only once during the preparation phase and then use the image processed in this way. The preparatory phase saves processor time.

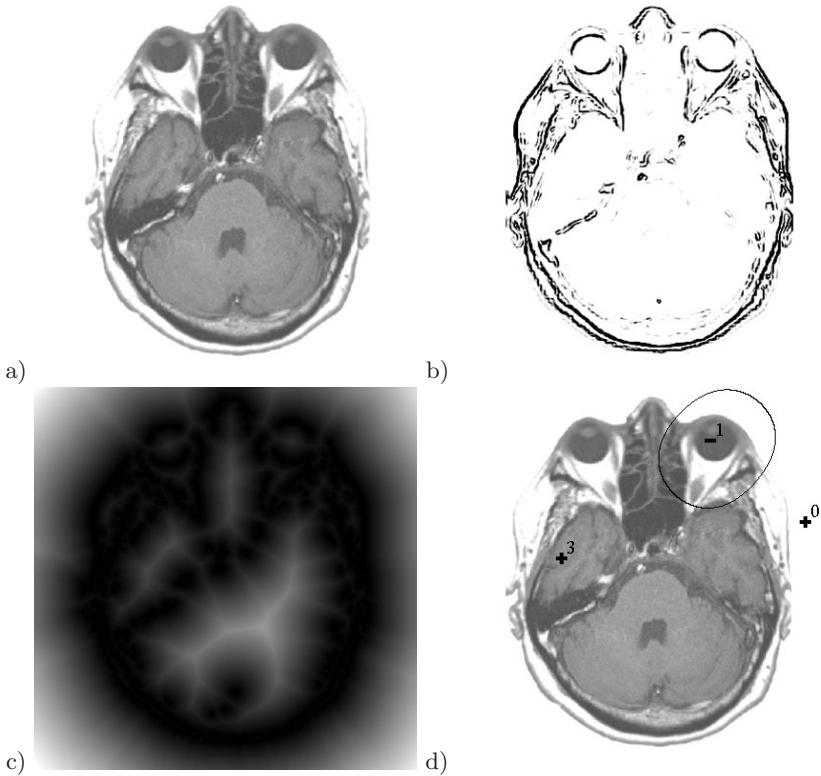


Fig. 3. (a) Cross-section of human skull. (b) Contours detected on the image of a skull. (c) Contour image after blurring. (d) Initial contour against a medical image.

3 Detection of an Eyeball

In the work, potential active contours in the form described in [8] have been applied. This experiment is based on a medical image showing a cross-section of human skull (Fig. 3a). The contours have been detected, as shown in Fig. 3b, which has subsequently been blurred (Fig. 3c). The method was initialized with one object source and four background sources, according to Fig. 3d. Method parameters are given in Table 1 – for their detailed description see [8]. The steps of the experiment are shown in Fig. 4. The experiment was repeated ten times,

Table 1. Adjusting of a contour to a human eyeball - parameters of the method

Simulated annealing	
Initial temperature T_0	-
The calculation of initial temperature T_0	YES
The number of iterations L needed to calculate initial temperature T_0	100
Maximum number of iterations M	3000
Temperature change interval L_T	30
Cooling schedule factor α	0.95
Normalization	NO
Markov process	YES
Markov chain length L_M	30
Move generator	
Position disturbance radius r	1.0
Charge disturbance factor γ	0.02
The probability of source number change p_z	0
Source chance equalizing	YES
Internal energy	
Desired area S_z	1300
Desired length L_z	-
Punishment for lack of consistency	NO
Internal energy weight w_{int}	4.0
Shape energy weight w_s	-

Table 2. Selected energy and temperature values for the process shown in Fig. 4

Initial temperature	3.04e10
Initial energy:	54342.0
Final temperature:	1.80e8
Final energy:	633.0

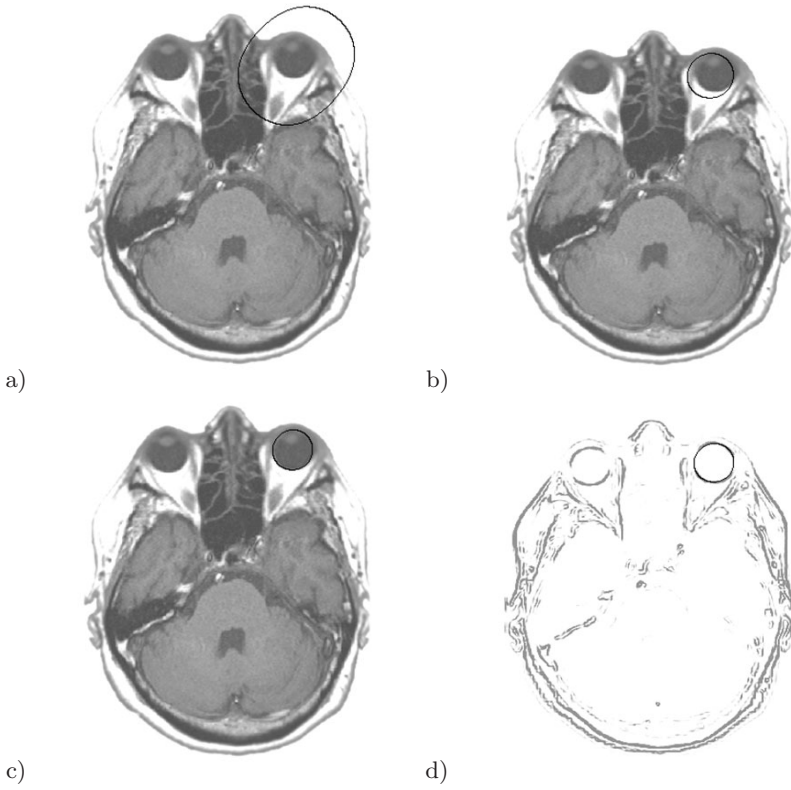


Fig. 4. The process of human eyeball retrieval: (a) Start, (b) 1000th iteration, (c) Final (3000th iteration), (d) Final contour against the contour image

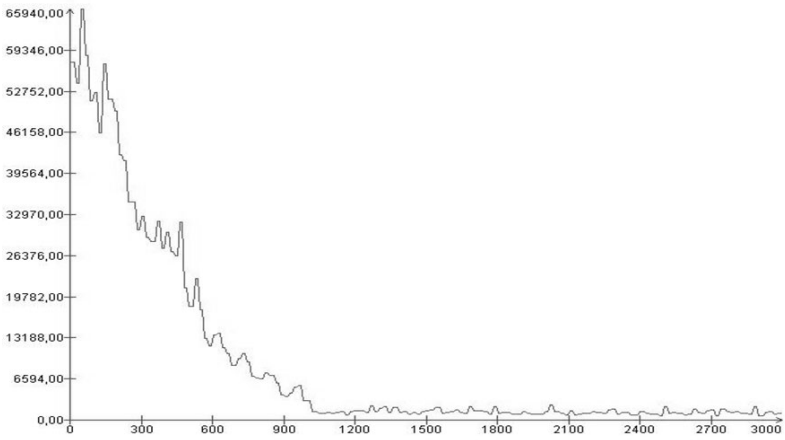


Fig. 5. Diagram of contour energy changes in function of number of steps in the algorithm

Table 3. Results of successive iterations. High standard deviation value is a result of two unusually high final energy values.

Test number	Final energy
1	642
2	647
3	1862
4	630
5	649
6	638
7	641
8	1853
9	631
10	649
Minimum	630
Maximum	1862
Average	944.5
Standard deviation	513.02

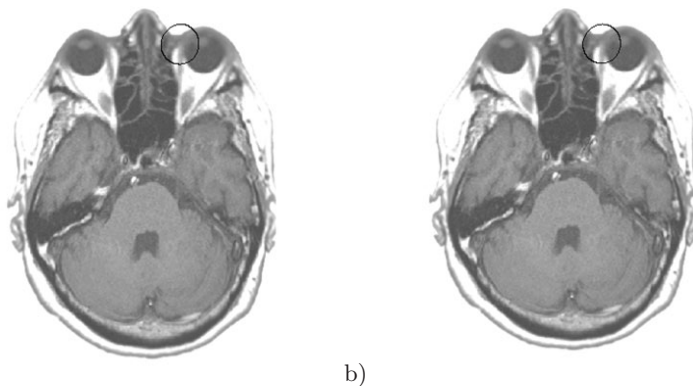


Fig. 6. The resultant contour against the image. (a) The final position of the contour for test 3 in Table 3 ($E = 1862$), (b) the final contour for test 8 ($E = 1853$).

in order to check the repeatability of the result - Table 3. Images Fig. 6a and Fig. 6b show unsuccessful attempts to adjust a contour to a particular fragment of an object. In these two cases, the final energy value is much different from those obtained in the other tests. So high a value is a result of an unsuccessful annealing process. As shown in Fig. 7, the method is able to find many objects simultaneously.

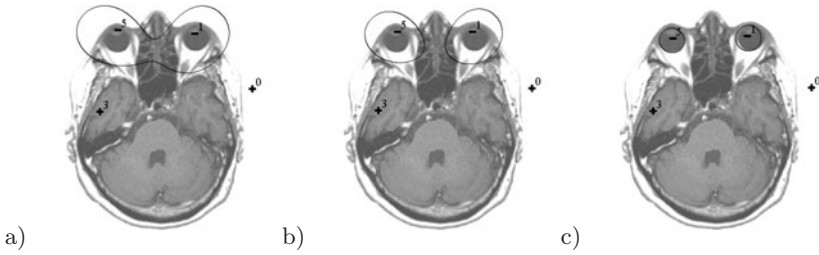


Fig. 7. Simultaneous detection of both eyeballs. (a) Initial contour ($E = 3902$). (b) The result after 100 iterations ($E = 2916$). (c) The result obtained after 2000 iterations ($E = 562$).

4 Summary

The experiments carried out in the present paper show how useful the potential active contour method can be for the segmentation of medical images. As proved by the majority of tests conducted here, the method is able to perform a correct segmentation of an eyeball. Both eyeballs have been detected on the image, although only one contour was used at the start. Moreover, the experiments are characterised by high repeatability of results. It is worth mentioning that a potential contour is based on a probabilistic optimisation algorithm. Therefore, the results depend to a large extent on a random factor. The algorithm has to be reiterated a few (or many) times and the best result is considered to be the final one. The method is more efficient if it is preceded by a preparation phase, during which the contours are found and the image is blurred.

References

1. Bankman, I.N. (ed.): Handbook of Medical Imaging, Processing and Analysis. Academic Press, San Diego (2000)
2. Gonzales, R.C., Woods, R.E.: Digital Image Processing. Prentice-Hall, Englewood Cliffs (2002)
3. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic Active Contours. *Int. Journal of Computer Vision* 22(1), 61–79 (2000)
4. Kass, M., Witkin, W., Terzopoulos, S.: Snakes: Active Contour Models. *Int. Journal of Computer Vision* 1(4), 321–333 (1988)
5. Tomczyk, A., Szczepaniak, P.S.: On the Relationship between Active Contours and Contextual Classification. In: Kurzynski, M., et al. (eds.) *Computer Recognition Systems. Proceedings of the 4th Int. Conference on Computer Recognition Systems - CORES 2005*, pp. 303–310. Springer, Heidelberg (2005)
6. Tomczyk, A.: Active Hypercontours and Contextual Classification. In: *Proceedings of the 5th International Conference on Intelligent Systems Design and Applications - ISDA 2005*, Wroclaw, Poland, pp. 256–261. IEEE Computer Society Press, Los Alamitos (2005)

7. Tomczyk, A., Szczepaniak, P.S.: Adaptive Potential Active Hypercontours. In: Rutkowski, L., Tadeusiewicz, R., Zadeh, L.A., Żurada, J.M. (eds.) ICAISC 2006. LNCS (LNAI), vol. 4029, pp. 692–701. Springer, Heidelberg (2006)
8. Tomczyk, A., Pieta, L., Szczepaniak, P.S.: Potential Active Contours - Basic Concepts, Mechanisms and Features. ASC, vol. 0047 (2008)

Potential Active Contours – Basic Concepts, Mechanisms and Features

A. Tomczyk¹, L. Pieta¹, and P.S. Szczepaniak^{1,2}

¹ Institute of Computer Science, Technical University of Lodz,
Wolczanska 215, 90-924, Lodz, Poland

² Systems Research Institute, Polish Academy of Sciences Newelska 6,
01-447 Warsaw, Poland
tomczyk@ics.p.lodz.pl

Summary. Potential active contours are based on the well known potential function method of classification, where the label assigned to the object depends on the distribution of other known and already classified objects. In the classic formulation of this approach, the known objects are fixed, while in the potential active contours' method their position and parameters are subject of optimization. In the present paper, the basic concept of potential active contours is described and its features are presented.

1 Introduction – Physical Background

Active contours can be interpreted as classifiers [1, 2, 3]. They can be applied in many diverse practical problems, for example image segmentation [4, 10, 11, 12] and classification of documents [5]. The characteristic feature of potential active contour is that it occurs in points with equal electrostatic field potentials. To be more specific, the contour is determined by a certain number of control points, which behave like electrically charged objects. The position of each control point s in a plane is specified by coordinates (x_s, y_s) and its charge by q_s . The sources with positive charge are called *the sources of the object*. Control points with negative charge are called *the sources of the background*. Each control point is a source of electrostatic field. The value of the potential function in each point of the field is:

$$V_s(p) = k \frac{q_s}{d} \quad (1)$$

where k is a constant and d is a Euclidean distance between the source and a given point of the field $p = (x_p, y_p)$:

$$d = d(p, s) = \sqrt{(x_s - x_p)^2 + (y_s - y_p)^2} \quad (2)$$

Since the field may result from more than one source, for calculating summary value of potential coming from all sources, the rule of superposition is used:

$$V(p) = \sum_{i=0}^n V_{s_i}(p) \quad (3)$$

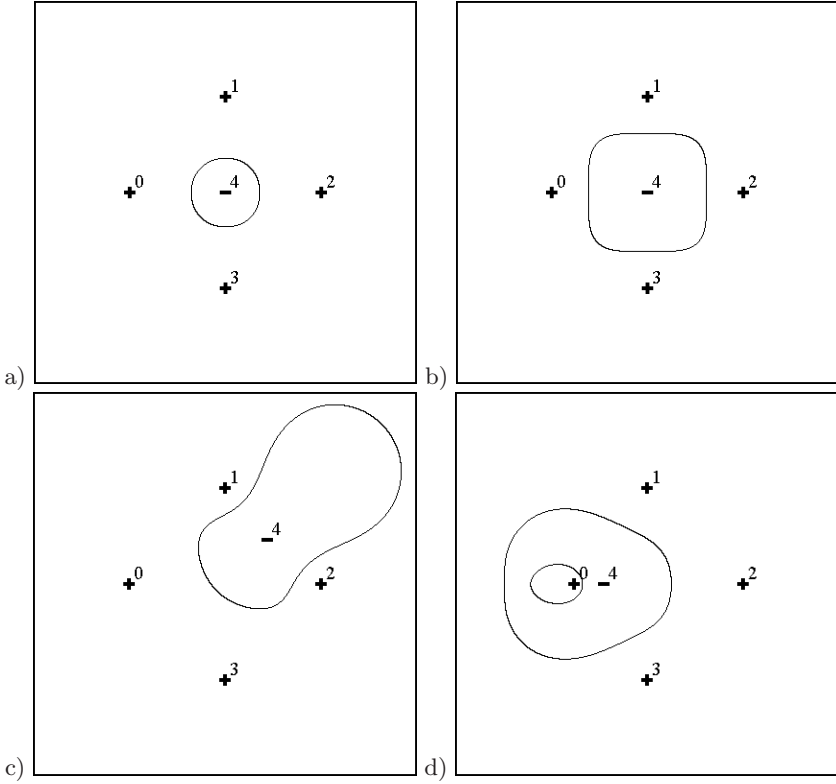


Fig. 1. Contours determined by four background's sources $q_{s_0} = q_{s_1} = q_{s_2} = q_{s_3} = 10$ and by one object's source q_{s_4} ; $S_T = \{s_0, s_1, s_2, s_3\}$, $S = \{s_4\}$: (a) $q_{s_4} = -15$, (b) $q_{s_4} = -30$, (c) new position of q_{s_4} , (d) new positions of q_{s_0} and q_{s_4}

Points in which scalar sum of all sources' potentials equals zero mark out contour's line that divides the image into inside and outside part of the object. Let S be a set of electric field sources:

$$S = \{s_0, s_1, s_2, \dots, s_n\} \quad (4)$$

Let the set of all object's sources be signified as S_O , and the set of background sources as S_T . Hence:

$$S = S_O \cup S_T \quad (5)$$

Potential contour is the following set of points:

$$K = \left\{ (x, y) \in \mathbb{R}^2 : k \sum_{i=0}^n \frac{q_{s_i}}{d((x_{s_i}, y_{s_i}), (x, y))} = 0 \right\} \quad (6)$$

Examples of potential contours are given in Fig. 1. As illustrated in the pictures, by using only five control points, one can obtain various figures (circle, approximation of a square, concave figure and a figure "full of holes"). By adding one

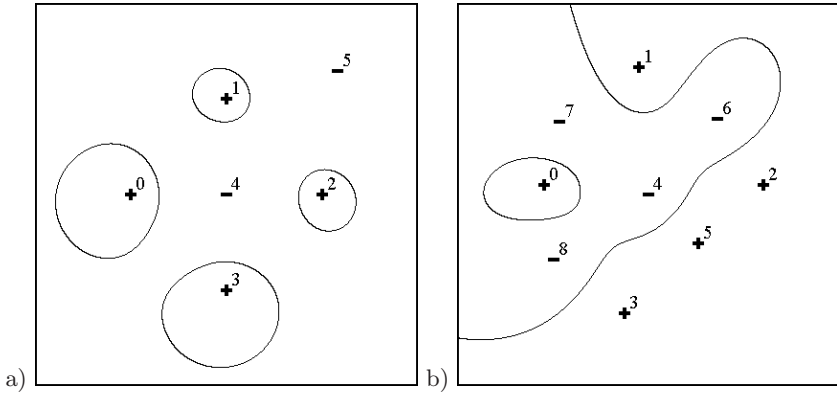


Fig. 2. Disjoint contours

point, one obtains a disconnected area - Fig. 2. It should be mentioned that it is possible to obtain a similar contour shape by different means, for example by appropriate choice of charge values or number of sources.

The description provided in the present paper is based on physical analogies. The idea of potential active contours can result from (and initially did result from) potential functions method that is one of the basic classification techniques. This approach was discussed in [4], which additionally presented also other potential functions.

2 Energy Function

Performing image segmentation, one aims at quality of image partition with respect to the homogeneity of the detected regions. Performing classification, one looks for the best classifier, i.e. the one that makes the smallest number of classification errors. In other words, one verifies the applied method with a performance index. In the active contours theory, the performance index that evaluates a given contour c is called *energy function*, denoted as $E = E(c)$. The total energy usually consists of two components, namely the external E_{ext} and the internal E_{int} :

$$E = w_{ext}E_{ext} + w_{int}E_{int} \quad (7)$$

In the external energy, the image is considered, while in the internal one, the form of the desired shape is reflected. Usually

$$E_{ext} = \sum_{(x,y) \in K_c} P(x,y) \quad (8)$$

where $P(x,y)$ is a function defined in the image. Frequently, one considers the intensity of light $I(x,y)$ at the pixels and one takes

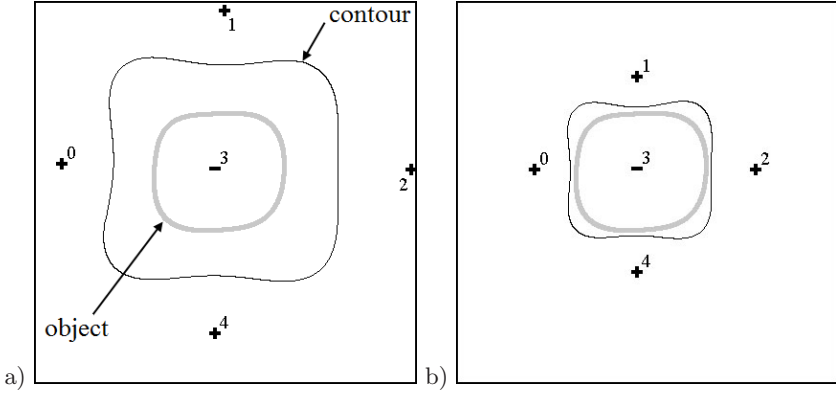


Fig. 3. Dependence between energy and the length of the contour: (a) $E_{ext} = 197370$, $E_{ext}^N = 255$, (b) $E_{ext} = 89505$, $E_{ext}^N = 255$

$$E_{ext} = \sum_{(x,y) \in K_c} I(x,y) \quad (9)$$

where K_c - the set of all points of contour c . The function I has a value between zero (black) and 255 (white). Another solution could be utilization of image gradient or other transformation of the image in P . Note that E_{ext} tends to zero when the contour approximates the desired object.

Since external energy defined in this way depends on the length of the contour (e.g. it becomes lower if the contour's perimeter gets smaller - cf. Fig. 3), a normalization mechanism was introduced. Owing to the mechanism, the energy value is not influenced by the length of the contour:

$$E_{ext}^N = \frac{1}{N} E_{ext} \quad (10)$$

where E_{ext}^N - normalized external energy, N - the number of pixels of the considered contour. External energy normalized in this way does not depend on the length of the contour, but on the average intensity of color at each pixel. In order to stop the contour from crossing the boundaries of the image, we can apply a punishment mechanism.

The internal energy reflects the contour's shape and does not depend on the image.

$$E_{int} = w_a E_a + w_l E_l + w_s E_s \quad (11)$$

where E_a - internal energy connected with the area belonging to the contour, E_l - energy connected with the length of the contour's curve, E_s - energy related to the shape of the contour, w_a , w_l , w_s - coefficients (weights). Taking the following notation: $A(c)$ - the area inside contour c , i.e. the number of the pixels belonging

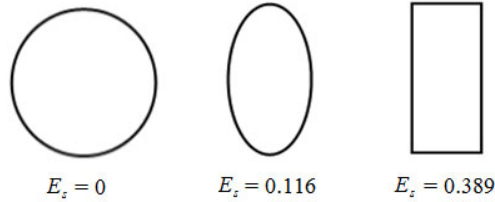


Fig. 4. Theoretical value of roundness coefficient for a few different shapes

to the inside part of the contour and $L(c)$ - the length of contour c defined as the number of pixels that constitute the contour's curve, one obtains:

$$E_a = E_a(c) = |A(c) - A_z| \quad (12)$$

$$E_l = E_l(c) = |L(c) - L_z| \quad (13)$$

$$E_s = E_s(c) = \frac{L(c)}{2\sqrt{\pi A(c)}} - 1 \quad (14)$$

where A_z and L_z denote desired area and desired length, respectively. E_s is the contour's roundness parameter (roundness coefficient). The closer the coefficient's value to zero, the more the contour's shape resembles a circle - cf. Fig. 4. There are many other shape coefficients, which usually:

- Should be able to differentiate various object shapes.
- Should be constant regardless of some image transformations, such as rotation, shift and scaling.
- Should possess the ability of being computed quickly, if they are to be used in the real-time system.

3 Simulated Annealing

Simulated annealing is a general probabilistic algorithm locating a good approximation of the global optimum of a given function in large search space. It was first introduced by S. Kirkpatrick, C. D. Gelatt and M.P. Vecchi in 1983 [6] as a continuation of Monte Carlo method.

The state of the system changes slightly with every step of the algorithm, as many potential solutions are taken into consideration. If new random solution is better, i.e. has lower energy value, then it is accepted and replaces the old one. The system can also move to a new worse state with a probability that depends on the difference between the corresponding energy function values and a parameter T called the temperature. During an annealing process, parameter T gradually goes to zero, which has a large influence on the evolution of the state. When the temperature is high, the probability of accepting a worse solution is also very high, which means that the solution move across the whole search

space randomly. The probability that a bad solution is chosen decreases with temperature reduction. Consequently, the search space of considered solutions becomes smaller and smaller, until the temperature is reduced to zero. Then, bad solutions are entirely ignored, and the system moves toward the nearest local minimum. The process of temperature reduction is called a "cooling schedule". The fact that the system is allowed to move to a new state, even if it is worse than the current one, prevents the method from becoming stuck in a local minimum.

Let c denote a contour (solution) from the set of all possible solutions C . Here, we consider the energy function $E = E(c)$ where $c \in C$ which is optimized. Every iteration considers a certain contour c' that is close to contour c . Thus, c' is called a neighbor of c . Considering the neighbors of the current state, the system probabilistically decides to move to a new contour or remain in the current state. Let $E = E(c)$ and $E' = E(c')$. Then the probability of a move to a new state c' is expressed by the formula:

$$p(c, c', T) = \begin{cases} 1 & \text{if } E' < E \\ e^{-\frac{E'-E}{T}} & \text{if } E' \geq E \end{cases} \quad (15)$$

Note that the probability of accepting a move becomes smaller, if the difference $\Delta E = E' - E$ increases, which means that moves to the states of higher energy are more likely when energy difference is smaller. The algorithm is completed, if the system reaches a sufficiently low energy state, or if there is no energy change (after a number of iterations), or after an assumed maximum number of iterations M . Additional mechanisms of the algorithm:

- The temperature is not reduced at every step of the algorithm. The temperature does not change with every iteration. The frequency of changes depends on the "temperature change interval" L_T . This parameter relates to the number of steps between subsequent temperature changes resulting from the cooling schedule. Thanks to that system is able to evolve through a given number of steps, with constant temperature.
- Markov chain. The Markov chain influences the annealing algorithm in that after a number of iterations L_M , the contour recorded as the best solution is accepted as the current solution. The period in which the best contour is remembered is the parameter of the method called "Markov chain length".
- Initial temperature T_0 can be calculated automatically in the initialization process. The method implements a heuristic that helps to calculate initial temperature, assuming that in the initial phase of the cooling process the changes toward a higher energy value should not exceed 80%. The temperature is calculated on the basis of contour energy changes caused by the move generator. The move generator makes the initial contour c_0 undergo numerous changes:

$$c_l = neighbor(c_{l-1}) \quad (16)$$

for $l = 1, \dots, L$. During those changes the difference $\Delta E_l = E(c_l) - E(c_{l-1})$ is recorded. Temperature T_0 is calculated according to the formula:

$$T_0 = -\frac{\Delta\bar{E}}{\ln 0.8} \quad (17)$$

where $\Delta\bar{E} = \frac{1}{L} \sum_{l=1}^L |\Delta E_l|$ is an average energy change. The length of initialization process L is "the number of iterations needed to calculate initial temperature".

To apply the simulated annealing method to any optimization task, one needs to specify:

- Space state and the representation of a state in space.
- Move generator - the method of choice of a new state, called *neighbor* above.
- Cooling schedule - the method of temperature change.
- Initial temperature parameter T_0 .

The above mentioned elements have a large influence on the method's efficiency. The parameters must be selected and matched carefully, otherwise the search for an optimal solution may be very time-consuming or unsuccessful (it may give a result which is not a local minimum). Unfortunately, there are no universal solutions every problem requires a specific solution. In some situations, the simulated cooling method may be impeded, because of mismatched parameters (they should be chosen after a number of experiments).

In the method of potential active contours, the state space is the space of parameters characterising potential contours. These parameters are the position and charge of each source. The neighbouring solution is chosen by means of the Gaussian movement generator. The temperature of the system is reduced according to the exponential cooling schedule $T_{l+1} = \alpha T_l$ where $\alpha \in (0, 1)$ and the initial temperature is calculated automatically.

4 Move Generator

At every step of the cooling process, the move generator changes slightly the current contour, thus creating a new one. This change consists in the modification of the contour's sources. What changes is the position of one source or its charge. This change has a character of a probability standard normal distribution. During the change of charge, it is impossible to replace the source of object with the source of background, and the other way round. The generator possesses two additional mechanisms, namely the equalization of chances and the change of the number of sources.

4.1 Gaussian Move Generator

The normal distribution, also called the Gaussian distribution, is one of the most important probability distributions. It is defined by density function:

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (18)$$

where μ - means value of the distribution (expected value) and σ - standard deviation. If $\mu = 0$ and $\sigma = 1$, the distribution is called a standard normal distribution, and the density function is defined by the equation:

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{19}$$

The density function in the normal distribution is symmetrical about straight line $x = \mu$, achieving maximal value in point $x = \mu$. The shape of the function depends on parameters μ and σ .

The move generator uses the standard normal distribution (the Gaussian distribution with parameters $\mu = 1$ and $\sigma = 1$) for changing the source's position or its charge. Thanks to the normal distribution, small changes occur more frequently than the big ones. The big changes are possible, but very unlikely. The move generator makes a change of position or charge value with equal probability. At a single step of the simulated cooling algorithm, the movement generator operates as follows:

1. Select source s_i from among all available sources S with equal linear probability p_w .
2. With equal probability, choose whether to disturb the charge or the position of the selected source s_i .
3. If the change of source position was chosen, disturb the coordinates (x_{s_i}, y_{s_i}) of the selected source s_i .
4. If the change of charge was chosen, disturb the charge q_{s_i} .

The position of source s_i specified by coordinates (x_{s_i}, y_{s_i}) is disturbed as follows:

$$x'_{s_i} = x_{s_i} + \Delta(r) \quad \text{and} \quad y'_{s_i} = y_{s_i} + \Delta(r) \tag{20}$$

where $\Delta(r) = G(0, 1)r$ and $G(0, 1)$ random number from standard Gaussian distribution ($\mu = 0$ and $\sigma = 1$), r - constant scale, here called "position disturbance radius". The value of $\Delta(r)$ can be positive or negative, both cases equally likely. According to the definition of normal distribution:

- 68% of value $\Delta(r)$ lies inside interval $[-r, r]$
- 95.5% of the value lies inside interval $[-2r, 2r]$
- 99.7% lies inside interval $[-r, r]$

The change of charge q_{s_i} of source s_i is relative, that is:

$$q'_{s_i} = q_{s_i} + \gamma G(0, 1)q_{s_i} = q_{s_i}(1 + \gamma G(0, 1)) \tag{21}$$

where γ - constant movement generator parameter, here called "charge disturbance factor".

4.2 Equalization of Chances

The move generator described above treats all sources in the same way - each of them has equal chances to be chosen. If a contour consists of only one object

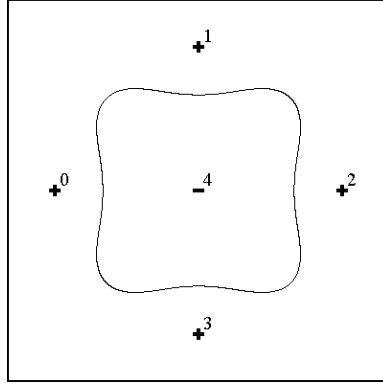


Fig. 5. Five sources: four background sources ($S_T = \{s_0, s_1, s_2, s_3\}$) and one object source ($S_O = \{s_4\}$)

source and four background sources (Fig. 5), each source is equally likely to be chosen $p_w(s_i)$. Note that it is the object's source $s_4 \in S_O$ that most influences the contour. It is rarely modified (statistically once in every five disturbances in the contour). To avoid a situation in which sources that have a major influence on a contour's shape are modified rarely the mechanism of equalization of chances was introduced. Owing to this mechanism, with every action of the movement generator, an object's source is modified as frequently as background's source. Thus, the move generator operates according to the following steps:

1. With equal probability, choose the type of source to be disturbed, i.e. object or background.
2. Select source s_i form the set S_T or S_O , depending on the probabilistic decision taken in step 1.
3. With equal linear probability choose whether to disturb the charge or position of the selected source s_i .
4. Depending on the decision made in step 3, disturb the charge or position of the selected source s_i .

This mechanism influences the probability p_w . In the aforementioned case:

$$p_w(s_i) = \begin{cases} 0.5 & \text{for } i = 4 \\ 0.125 & \text{for } i \in \{0, 1, 2, 3\} \end{cases} \quad (22)$$

4.3 Modification of the Number of Sources

The move generator is also capable of modifying the contour by increasing and reducing the number of sources of the potential contour.

It is decided at random with equal probability whether a source is to be added or delete. The operation of source number reduction consists of random choice of the source $s_i \in S$, which is next deleted from the set of all sources $S' = S \setminus \{s_i\}$.

If we want to increase the number of sources, we add a new source s_{n+1} to the system ($S' = S \cup \{s_{n+1}\}$), hence:

- The position of the source is random, i.e.: $x_{s_{n+1}}$ and $y_{s_{n+1}}$ are chosen using linear probability distribution.
- The charge of the source is chosen randomly with a standard normal distribution, i.e. $q_{s_{n+1}} = G(0, 1)$.

The frequency of source number change is determined by the parameter p_z , called "the probability of source number change".

4.4 Move Generator – Suggestions for Further Research

Below, the author presents his ideas and suggestions that were not implemented in the present work, but could largely improve the efficiency of move generator:

- Sources may be modified with a probability that depends on the influence they have on the contour's shape. The smaller the influence, the fewer chances the source has to be modified. The influence of sources on the contour's shape can be calculated on the basis of their number, type, charge, their distance from the contour, or the potential field distribution.
- The move generator should be able to modify or add more than one source in a single step of simulated cooling algorithm. In some situations (when only one source is modified), a contour achieves a state, in which finding a better solution is impossible. Only a simultaneous change of a few sources makes it possible to move from the local minimum toward a better solution [4].
- Source adding may be performed in an adaptive way [4]. The sources should not be placed in random positions, otherwise they are very unlikely to improve the system. The position of a new source should be precisely specified so as to improve the system, thus enabling the contour best possible adjustment to complicated shapes.
- Instead of choosing a source, the movement generator can choose any pixel that will next disturb the position or charge of contour's sources, proportionally to the distance between that pixel and the source. Such an expansion of application might be considered as an equivalent to constraints described in [8]. It would also allow the user to include expert knowledge.

5 Summary

Potential contour, as every active contour method [7, 8], replaces performs segmentation by optimization, i.e. it searches for an optimal contour characterized by a minimal possible energy value (which depends on the shape and background of the contour). Additionally, the optimization process is controlled by a stochastic algorithm of simulated annealing [6], in which the temperature is reduced exponentially. The contour is disturbed by means of a Gaussian move generator, with the possibility to add or delete control points. The point are sources of electrical field and represent the contour. Moreover, the energy of the

contour does not depend only on the pixels' brightness, but also on such shape parameters as perimeter, area and roundness coefficient.

Potential active contour can be identified with the task of classification [1, 2], in which pixels are divided into two groups. The former is connected with the desired object on the image, the latter with the background (regarded as redundant information). The aim of the method is to specify a group of pixels, which will enable the user to find the contours of a specific object or a group of objects, as in the case of medical image analysis [9].

References

1. Tomczyk, A.: Active Hypercontours and Contextual Classification. In: Kwasnicka, H., Paprzycki, M. (eds.) Proceedings of the 5th International Conference on Intelligent Systems Design and Applications - ISDA 2005, Wroclaw, Poland, pp. 256–261. IEEE Computer Society Press, Los Alamitos (2005)
2. Tomczyk, A., Szczepaniak, P.S.: On the Relationship between Active Contours and Contextual Classification. In: Kurzynski, M., Wozniak, M., Puchala, E., Zolnierek, A. (eds.) Proceedings of the 4th International Conference on Computer Recognition Systems (CORES), pp. 303–311. Springer, Heidelberg (2005)
3. Tomczyk, A., Szczepaniak, P.S.: Adaptive Potential Active Hypercontours. In: Rutkowski, L., Tadeusiewicz, R., Zadeh, L.A., Żurada, J.M. (eds.) ICAISC 2006. LNCS (LNAI), vol. 4029, pp. 692–701. Springer, Heidelberg (2006)
4. Tomczyk, A.: Image Segmentation using Adaptive Potential Active Contours. In: Kurzynski, M., Puchala, E., Wozniak, M., Zolnierek, A. (eds.) Computer Recognition Systems (CORES), Wroclaw, Poland (Series: Advances in Soft Computing), pp. 148–155. Springer, Heidelberg (2007)
5. Szczepaniak, P.S., Tomczyk, A., Pryczek, M.: upervised Web Document Classification using Discrete Transforms, Active Hypercontours and Expert Knowledge. In: Zhong, N., Liu, J., Yao, Y., Wu, J., Lu, S., Li, K. (eds.) Web Intelligence Meets Brain Informatics. LNCS (LNAI), vol. 4845, pp. 305–323. Springer, Heidelberg (2007)
6. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by Simulated Annealing. *Science* 220(4598), 671–680 (1983)
7. Bakos, M.: Active Contours and their Utilization at Image Segmentation. In: Proceedings of 5th Slovakian-Hungarian Join Symposium on Applied Machine Intelligence and Informatics, Poprad, Slovakia, January 25-26, 2007, pp. 313–317 (2007)
8. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active Contour Models. *International Journal of Computer Vision* 1(4), 321–331 (1987)
9. Pieta, L., Tomczyk, A., Szczepaniak, P.S.: Medical Image Analysis using Potential Active Contours. ASC, vol. 0047. Springer, Heidelberg (2008)
10. Gonzalez, R., Woods, R.: Digital Image Processing. Prentice-Hall Inc., New Jersey (2002)
11. Sonka, M., Hlavac, V., Boyle, R.: Image Processing, Analysis and Machine Vision. Chapman and Hall, Cambridge (1994)
12. Yong, Y., Chongxun, Z., Pan, L.: A Novel Statistical Approach for Segmentation of Single-Channel Brain MRI using an Improved EM Algorithm. *Journal of Applied Computer Science* 13(1), 113–125 (2005)

Fractal Magnification of Medical Images

Jan Kwiatkowski and Wojciech Walczak

Faculty of Computer Science and Management, Wrocław University of Technology
jan.kwiatkowski@pwr.wroc.pl, wojciech.walczak@student.pwr.wroc.pl

Summary. Image magnification takes important part during medical examination. It is especially very helpful in observing small details and their manual measurements. There are a number of magnification methods, however most of them cause image distortions. The paper presents a fractal magnification method, which allows minimizing these negative effects besides the blocks effect.

1 Introduction

In the last few years, the quickest development of non-invasive medical diagnostic methods has been observed. Depending on the used diagnostic instrument, different format of medical images are used but all of the images are characterized by a very limited image size, for example, Computerized Tomography images are not larger than 512×512 pixels, Magnetic Resonance images than 256×256 , etc. Therefore, image magnification takes important part during medical examination; it is especially very helpful in observing small details and their manual measurements. For example, the B-mode ultrasonography allows visualizing the lumen and walls of arteries. It can be used to detect atherosclerosis in the earliest stages and study its progression or regression. Most often, computerized analyzing systems with manual tracking of the echo interfaces for measurement of Intima-Media Thickness (IMT) are used. Without image magnification, it is hard and sometimes impossible to mark the interfaces with high precision.

There are a number of image magnification methods. Nevertheless, most of them cause image distortions resulting in loss of sharpness of edges in the image. Image blurring, pixelization, and blocks effect are the most unwelcome distortions. The solution can be using fractal magnification that allows minimizing most of these negative effects, besides the blocks effect. However, fractal magnification is a very time consuming operation. The paper deals with the use of fractal magnification for medical images. The short description of developed algorithm and the way of its parallelization are presented.

The structure of the paper is following: section 2 describes the fundamentals of fractal magnification. The algorithm and its parallelization scheme are described in section 3. The results of experiments are presented in section 4. Finally, section 5 concludes received results and discusses ongoing work.

2 Fundamentals of Fractal Magnification

Fractal magnification is a process consisting of fractal encoding of the original image and fractal decoding to size larger than original. The encoding produces a fractal operator W , which is a set of contractive affine transformations: $W = \sum_{i=1}^n w_i$. The transformations map parts of the image (domain blocks D_i) into range blocks R_i . The range blocks are disjoint and cover entire image. Besides, the range blocks are usually two-times smaller than the domain blocks. Thus, the goal of the encoding process is to find pairs of range blocks and domain blocks. The error (measured in RMS^2) between each range block and corresponding domain block (after transformations: spatial contraction, intensity transformation, symmetry operation) shall be smaller than a preset fixed value. If the image is fractally magnified, the operator W does not have to be stored to file – it is instantly decoded directly from the description of the fractal operator stored in memory.

The operator W defines an unequivocal fixed point – attractor (A). The attractor can be reached from any starting image f_0 through iterations of the operator W . The iterations produce successive approximations of the attractor. Usually from 7 to 10 iterations are performed during decoding. When the size of the initial image f_0 is larger than the size of the original image f , the image is fractally magnified – the positions and sizes of the range and domain blocks have to be scaled. This is possible because the “similarity” of the range and domain blocks for the given image is independent of the magnification scale and their mutual positions are relative inside the image. Enlarged range and domain blocks result in generation of additional image details during decoding. However, the decoded (magnified) images are not free from deformation – blocks effect and arising artifacts reduce their quality.

3 Proposed Magnification Method

Although all fractal coding methods are based on the same fundamentals, plenty of design decisions have to be made to develop a complete coding method. Most efforts were made on preserving the fidelity of the magnified because this is the most important aspect in medical imaging (section 3.1). Attempts to reduce the encoding time also were made (section 3.2).

3.1 Image Fidelity

In order to assure high image fidelity several aspects are considered. Image partitioning and blocks’ splitting techniques are optimally chosen as well as the error caused by quantization during encoding is eliminated.

The image may be partitioned into range blocks in various ways. The choice of the partitioning method has a great influence on the fidelity of the magnified image. The best results can be obtained when the partitioning of the image adapts to the content of the image. Such partitioning methods can be divided into two groups: hierarchical and split-and-merge schemes.

The hierarchical partitioning methods begin with some initial partitioning of the image, e.g. uniform partitioning. In the implemented algorithm, there is only one range block, which covers the entire image. Then for each range block, the best matching domain block is searched. If the distance between a range block and found domain block (measured in mean square error) is larger than some value (constant for the entire encoding process) then the range block is divided into several range blocks. The number and shape of the newly created range blocks vary among different hierarchical methods. The split-and-merge approaches divide the image into partitions in two-phase process that ends before the searching for transformations is started. Only the partitioning schemes based on right-angled blocks are considered because others, based on triangles or polygons [3], are inferior. The irregular regions (split-and-merge approach) are the best option for fractal compression when the image is compressed with compression ratio higher than 10 : 1. However, the rate-distortion characteristic is irrelevant in fractal magnification. After scrutinizing the results of best hierarchical methods – based on horizontal-vertical partitioning (HV) [1, 2] – and best method based on irregular regions [4], it is assumed that the HV approach allows obtaining better image fidelity but with higher encoding time cost. In case of medical images, the image fidelity is superior because it can affect the diagnosis. Thus, the HV partitioning scheme has been chosen for implementation as the one that potentially can give the most accurate images not only when they are decoded to the same size as original but also when they are magnified.

In the implemented method, the range block is split into two when there cannot be found matching domain block. The cutting line is horizontal or vertical and its location depends on the content of the block. The most significant edge in becomes the cutting line [1]. However, the implemented edge detection mechanism uses modified formulas to find the most significant edge:

$$v_m = g(m) \cdot \left| \sum_n^{N-1} r_{m,n} - \sum_n^{N-1} r_{m+1,n} \right| \quad h_n = g(n) \cdot \left| \sum_m^{M-1} r_{m,n} - \sum_m^{M-1} r_{m,n+1} \right|$$

where $r_{m,n}$ is the pixel value of the range block at coordinates (m, n) , and the size of the range block is $M \times N$. If v_m is higher than h_n then the block is divided vertically after the m^{th} column. Otherwise, the block is divided horizontally after the n^{th} row. The modified v_m and h_n formulas ensure that the most significant edge will be not missed – the original formulas detect only the edges where a row or column with higher mean intensity is prior during traversing the block. The function g prevents the rectangle blocks from degradation to a line or a point. The g is the same as in the work of Fisher [1]: $g(x) = \min(x, X - 1 - x)/X$, where X is the number of rows (v_m) or columns (h_n).

Other approach locates the cutting line in a position that will minimize the sum of pixel intensity variances of the produced blocks [2]. Experiments performed on ultrasonographic images show that the modified edge detection algorithm results in better image fidelity [5]. Thus, the modified edge detection method is used for implementation.

The decoding process also has an influence on the image fidelity. The attractor approximations created in successive iterations of the decoding algorithm are normally stored using raster images, i.e. on two-dimensional arrays of integers. Quantization of the pixel values causes propagation of error to the result of consecutive decoding iterations. The propagation of quantization error may cause difficulties with reaching the correct values of brightness of some pixels in the final decoded image. *Accurate Decoding with Single-time Quantization* stores these approximations on two-dimensional arrays of real numbers. The pixel values do not have to be quantized after each iteration, the quantization is just after the last iteration. [6]

3.2 Time Cost Reduction

Optimization of the codebook, variance-based acceleration and parallelization are introduced to the algorithm in order to reduce the encoding time.

During searching for matching range blocks and domain blocks, each range block is compared with spatially contracted domain blocks, which compose the virtual codebook. The implemented algorithm does not contract domain blocks independently. Before starting the encoding, the whole image is contracted and the codebook is determined directly on the contracted copy of the image. This speeds up the encoding because domain blocks may overlap.

The implemented method uses global codebook, i.e. the codebook is determined at the beginning of the encoding and it is the same for all range blocks. The encoding time can be reduced when local codebook is used, where the density of codebook blocks decreases with spatial distance to currently considered range block or the codebook does not contain distant codebook blocks at all. However, when ultrasonographic images are fractally encoded, the probability of finding matching domain block does not increase when the spatial distance between the domain and range blocks decreases [5]. Thus, the local codebook would cause to high distortions.

The HV partitioning is characterized with large variety of sizes of the range blocks. For each possible size of range blocks, codebook blocks have to be included to the codebook. Although the pixel values of any codebook blocks can be taken from the contracted image, the descriptions of the codebook blocks have to be stored in the memory. The descriptions include the size, location of the blocks, and symmetry operation identifier, but also inner products, which are used to calculate the optimal intensity transformation coefficients for each pair of range and domain block. The calculations of the inner products, which are the most time consuming, may be postponed to the first access to a codebook block during the search for transformations. This reduces the encoding time because the inner products will never be calculated for the codebook blocks that will not be compared with any range block.

When the codebook is very numerous (large image to encode and low domain offset) then a very large amount of memory is required to store the descriptions.

This memory can be saved by a proposed version of the codebook – on-the-fly codebook. The codebook does store no descriptions of the codebook blocks, they are calculated when a codebook block is accessed. This solution may cause additional time costs because the inner products for some codebook blocks will have to be recalculated several times (as many times as many range blocks have the same size like the codebook block). There is also proposed a hybrid solution – the descriptions are only stored for the smallest codebook blocks because most likely more than one range block will be compared during the search with these codebook blocks.

The encoding time can be reduced by increasing the offset between domain blocks. However, this simple solution eliminates domain blocks independently of the blocks' content and significantly reduces the image fidelity. The implemented method uses a variance-based acceleration technique inspired by the work of He, Yang and Huang [7]. Only the domain blocks that are least likely to be paired with currently considered range block are eliminated.

For all range and codebook blocks, intensity variances are calculated from the equation: $\sigma(B) = \frac{1}{n} \left[\sum_{i=0}^{n-1} b_i^2 - \left(\sum_{i=0}^{n-1} b_i \right)^2 \right]$, where $B = b_0, \dots, b_{n-1}$ is the measured block and n is the number of pixels B . The error between the range and codebook blocks is very close related with the distance between the variances of the blocks [8]. During the searching for matching codebook blocks to each remaining range block, the codebook blocks with variance lower than the variance of the range block or higher than $\sigma(R_i) + \Delta\sigma$ are omitted. The $\Delta\sigma$ is set before starting the encoding. Range blocks with variance lower than a given (preset) value are treated as shade blocks, i.e. they are approximated with a fixed block. Thus, further speedup is attained because, for these range blocks, no search after matching codebook blocks is required.

The proposed parallelization scheme uses a fixed number of threads to find the affine transformations that compose the fractal operator. The encoding acceleration through parallel processing is very promising because the transformations can be found independently.

The threads share a queue of range blocks that wait to be encoded. Each thread consumes one range block from the beginning of the queue and performs the entire encoding process, i.e. determines the codebook for the range block, finds the best matching codebook block, computes the intensity transformation parameters and the distance between the range and codebook blocks. If the error between the range block and the best matching codebook does not fulfill the tolerance criterion (TC, expressed in RMS^2) then range block is split within the same thread and the newly created range blocks are added to the common queue. However, if the error between the range block and the transformed codebook block meets the tolerance criterion then the description of the affine transformation is added to the thread's internal data structure. When the queue with range blocks to encode is empty and there is no active thread that can produce new range blocks, the threads send the descriptions of found transformations to the main thread and end their life.

4 Experimental Results

The two most important aspects in fractal magnification of medical images were assessed through experiments: protecting the information stored in the image and minimizing the time needed for magnification.

Objective measures are used to assess the fidelity of the magnified images. Peak signal to noise ratio (*PSNR*), expressed in decibels (*dB*), and mean square error (*MSE*) are most often used:

$$PSNR = 10 \cdot \log_{10} \left(\frac{\max_X^2}{MSE} \right)$$

$$MSE = \frac{1}{M \cdot N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left(X(m, n) - \tilde{X}(m, n) \right)^2$$

where X is the original image, \tilde{X} – the magnified image fractally demagnified to the original size, M – number of pixels in a row, N – number of pixels in a column, $\max_X = 2^b - 1$ – the maximal possible pixel value of the image X (b – bits per pixel). Other measures, like image fidelity (IF), average error (AE), mean absolute error (MAE), maximal error (ME) are also calculated in order to gain additional information about the distortions caused by magnification. The magnified image has to be rescaled to the original size before the measurements are performed because the objective measures can be used only for same-sized images.

Table 1. Comparison of fractal magnification and bicubic interpolation

method	PSNR	MSE	IF	ME	MAE
fractal (optimal TC)	37.77	12.17	0.9982	42.00	1.78
fractal (TC = 1)	37.58	12.68	0.9982	42.23	1.82
bicubic	36.58	16.69	0.9978	22.4	2.57

The results of the experiments are summarized in the table 1, where the average values for forty images can be found. There are two different fractal magnification results presented. The first row informs about results of measurement when each image is encoded with optimal tolerance criterion value. Experiments show, that these optimal TC are most often within range (0.4, 2.5). There cannot be found a TC value that would result in optimal fidelity of any magnified image. However, the value 1 assures that the fidelity of the magnified image will be close to maximal.

The distribution of the pixels errors is different in the methods. Bicubic interpolation is used as a reference standard because it produces image of high

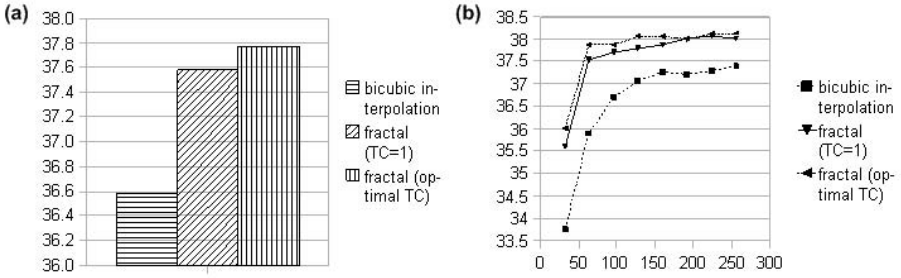


Fig. 1. Results of PSNR measurements for fractal magnification and bicubic interpolation. (a) – average values for all images, (b) – average values for different sizes of original images.

quality and is commonly used. Although the overall error caused by the fractal magnification is lower than the one caused by bicubic interpolation, there can be observed rare pixels where the errors are higher. The interpolation decreases the brightness of the image during magnification, while the fractal magnification most often increases the brightness.

The fidelity of the magnified image depends also on the size of the original image. When the image is smaller, the virtual codebook contains less blocks and it is harder to find a good match between range blocks and codebook blocks. Thus, the larger an image is, the better quality of the magnified image can be obtained. The quality of the image is significantly lower when images 32×32 are magnified.

The fidelity, quality and usefulness of images magnified with fractal method are higher than with bicubic interpolation according to subjective assessment made with cooperation with medicine physicians. Visibility and fidelity of small details, readability of edges in the image and lack of image blurring are the main reasons that state the superiority of fractal magnification. The quality of the fractally magnified image is lowered by blocks effect but it does not cause problems with reading the image.

A 256×256 image can be magnified in about 210 seconds, 192×192 image in 60 seconds, 128×128 in 10 seconds, and 64×64 in less than one second (tests performed on machine with 2 GHz processor and 1 GB RAM). The use of shade blocks can reduce the encoding time by 5% with almost no cost in image fidelity. Further improvement entails loss of image quality – 30% time reduction gives PSNR lower by about 15%. Reduction of the codebook by introducing limit of variance distance between range blocks and codebook blocks gives further acceleration. The encoding time is decreased by 30% when the PSNR is reduced by 0.6%. The cost of 4% gives two times shorter encoding time. Parallelization gives in almost ideal results – use of pool of two threads, working on two-core processor, reduces the encoding time about 1.9 times.

5 Conclusions and Future Work

Received results for the developed algorithm confirm the usefulness of the fractal magnification in medical diagnosis. All objective measures as well as subjective assessment of magnified images indicate the superiority of fractal magnification over bicubic interpolation. Image blurring is very visible when interpolation methods are used. In fractal magnification, artifacts as image blurring or pixelization do not occur. The distortions are caused only by the block effect. Nevertheless, this type of distortions can be reduced by introducing postprocessing [9] and it will be the subject of the future work.

The experiments were performed only on ultrasonographic images. Although, similar conclusions are expected for other types of medical images, several of the remaining types shall be investigated during further research.

References

1. Fisher, Y., Menlove, S.: Fractal encoding with HV partitions. In: Fisher, Y. (ed.) *Fractal Image Compression: Theory and Application*, Springer, Heidelberg (1994)
2. Saupe, D., Ruhl, M., Hamzaoui, R., Grandi, L., Marini, D.: Optimal hierarchical partitions for fractal image compression. In: *Proc. ICIP 1998 IEEE International Conf. on Image Proc.*, Chicago (1998)
3. Wohlberg, B., de Jager, G.: A review of the fractal image coding literature. *IEEE Trans. on Image Proc.* 8(12), 1716–1729 (1999)
4. Ochotta, T., Saupe, D.: Edge-based partition coding for fractal image compression. *The Arabian Journal for Science and Engineering, Special Issue on Fractal and Wavelet Methods* 29(2C) (2004)
5. Walczak, W.: *Fractal compression of medical images*. MSc Thesis, Wrocław University of Technology (2008)
6. Kwiatkowski, J., Kwiatkowska, W., Kawa, K., Kania, P.: Using fractal coding in medical image magnification. In: Wyrzykowski, R., Dongarra, J., Paprzycki, M., Waśniewski, J. (eds.) *PPAM 2001. LNCS*, vol. 2328, pp. 517–525. Springer, Heidelberg (2002)
7. He, C., Yang, S.X., Huang, X.: Variance-based accelerating scheme for fractal image encoding. *Electronics Letters* 40(2), 115–116 (2004)
8. Lee, C.K., Lee, W.K.: Fast fractal image block coding based on local variances. *IEEE Trans. on Image Proc.* 7(6), 888–891 (1998)
9. Zakhor, A.: Iterative procedures for reduction of blocking effects in transform image coding. *IEEE Trans. on Circuits and Systems for Video Technology* 2(1), 91–95 (1992)

Fuzzy Clustering in Segmentation of Abdominal Structures Based on CT Studies

Wojciech Więclawek¹ and Ewa Piętka²

¹ Silesian University of Technology, Faculty of Automatic Control,
Electronics and Computer Science, Institute of Electronics,
Department of Biomedical Electronics
wojciech.wieclawek@polsl.pl

² ewa.pietka@polsl.pl

Summary. In the current study a modification of the live-wire approach to image segmentation has been developed. A Fuzzy C-Means (FCM) clustering procedure has been implemented before the cost map function is defined. This shrinks the area to be searched resulting in a significant reduction of the computational complexity. The method has been employed to Computed Tomography (CT) studies. Segmentation of the abdomen structures has been performed in order to evaluate the method.

1 Introduction

Computer assisted radiological diagnosis requires often an analysis of individual anatomical structures, location of pathological regions, extraction of various features describing the texture and size. Image segmentation playing a vital role in numerous applications, relies on extraction of image homogeneous and non-overlapping regions, connected with objects in an image. The delineation of an anatomical structures requires often a problem dependent approach and makes one universal method not applicable.

In many cases the analysis does not yield correct results, and only an interactive approach may result in a successful edge delineation. In the current study a Live-Wire approach has been tested and modified in order to grand a fast and reliable method.

In Section 2 the traditional version has been presented. Its modification, which employs the Fuzzy C-Means procedure is described in Section 3. Results discussion conclude the paper.

2 Traditional Live-Wire Algorithm

The Live-Wire approach [1, 2, 3, 4] is based on image boundary analysis and consist of several stages. First, the computation of an image cost map is required. It reflects the edge features in each image pixel. Then, the image is represented

by a graph, in which each vertex corresponds to an image pixel and edges link the vertex with each of its eight neighbouring vertices. Each vertex refers to an element of the image cost map. Next, optimal paths are found in a graph. To choose the desired edge, the user specifies characteristic points, namely a seed point and a free point. By repeating the operation the segmentation of the given structure is made. A closed contour stops the segmented structure, and yields the final result.

2.1 Local Cost Map Definition

Because the Live-Wire algorithm belongs to the edge-based segmentation techniques their cost matrix is generated as a function of the image gradient magnitude and gradient orientation. The cost function components frequently employ [1, 2, 3, 4]:

- Laplacian zero-crossing – $f_Z(\mathbf{q})$
- Gradient magnitude – $f_G(\mathbf{q})$
- Gradient direction – $f_D(\mathbf{p}, \mathbf{q})$

where \mathbf{p}, \mathbf{q} denote pixels, described by two elements (x and y coordinate) vectors. The gradient components are found by implementing various mask sizes in a multiscale edge detection.

The final cost function is defined as a weighted sum of these three normalized components, i.e.:

$$l(\mathbf{p}, \mathbf{q}) = \omega_Z \cdot f_Z(\mathbf{q}) + \omega_G \cdot f_G(\mathbf{q}) + \omega_D \cdot f_D(\mathbf{p}, \mathbf{q}) \quad (1)$$

where weights have to keep a condition:

$$\omega_Z + \omega_G + \omega_D = 1 \quad (2)$$

This preserves the normalization of the cost function. In particular individual cost matrix components are also normalized in this way that low local edge cost corresponds to pixels with strong edge features.

2.2 Graph Searching

Graph searching algorithm (Algorithm 2.1) analyzes each vertex corresponding to an image pixel and its neighborhood, starting from the seed point (\mathbf{s}) interactively admitted by the user. The algorithm uses the local cost matrix, (equation (1)) and the seed point (\mathbf{s}). As a result, a tree rooted at the seed point is obtained. The tree is built upon an array (\mathbf{k}) showing the minimal cost path leading to the seed point. In each iteration a single image pixel (\mathbf{p}) with its neighborhood $\mathbf{N}(\mathbf{p})$ is analyzed.

Algorithm (Algorithm 2.1) starts at the user-defined seed point (\mathbf{s}) and places this point on an initially empty sorted list \mathbf{L} (hence $\mathbf{L} \leftarrow \mathbf{L} \cup [\mathbf{s}, g(\mathbf{s})]$).

Algorithm 2.1. *Live-Wire*(\mathbf{l}, \mathbf{s})

```

%  $\mathbf{l}$  – local cost map
%  $\mathbf{s}$  – seed point
%  $\mathbf{L}$  – priority queue
%  $\mathbf{g}$  – global cost map
%  $\mathbf{e}$  – array of analyzed pixels
%  $\mathbf{k}$  – numerical representation of tree
%  $\mathbf{N}(\mathbf{p})$  – neighborhood of pixel  $\mathbf{p}$ 
%  $\mathbf{p}$  – actually analyzed image pixel
%  $\mathbf{x}$  – image pixel
%  $\mathbf{q}$  – single pixel from  $\mathbf{N}(\mathbf{p})$ 

```

Input: \mathbf{l}, \mathbf{s}

Data structures: $\mathbf{g}, \mathbf{e}, \mathbf{L}, \mathbf{N}(\mathbf{p})$

Output: \mathbf{k}

Main:

$g(\mathbf{s}) \leftarrow 0$

$\mathbf{L} \leftarrow \mathbf{L} \cup [\mathbf{s}, g(\mathbf{s})]$

while $\mathbf{L} \neq \{\}$

```

do {
  do {
    do {
       $[\mathbf{p}, g(\mathbf{p})] \leftarrow L[x_1, g(x_1)]$ 
      for each  $\mathbf{q} \in \mathbf{N}(\mathbf{p})$ 
        if  $e(\mathbf{q}) = 0$ 
          then {
             $g_{tmp}(\mathbf{q}) \leftarrow g(\mathbf{p}) + l(\mathbf{q}, \mathbf{x}) \cdot (1 \text{ or } \sqrt{2})$ 
            if  $\mathbf{q} \in \mathbf{L}$  and  $g_{tmp}(\mathbf{q}) < g(\mathbf{q})$ 
              then {
                 $\mathbf{L} \leftarrow \mathbf{L} \setminus [\mathbf{q}, g(\mathbf{q})]$ 
                 $g(\mathbf{q}) \leftarrow g_{tmp}(\mathbf{q})$ 
                 $k(\mathbf{q}) \leftarrow \mathbf{p}$ 
                 $\mathbf{L} \leftarrow \mathbf{L} \cup [\mathbf{q}, g(\mathbf{q})]$ 
                SORT( $\mathbf{L}$ )
              }
            if  $\mathbf{q} \notin \mathbf{L}$ 
              then {
                 $g(\mathbf{q}) \leftarrow g_{tmp}(\mathbf{q})$ 
                 $k(\mathbf{q}) \leftarrow \mathbf{p}$ 
                 $\mathbf{L} \leftarrow \mathbf{L} \cup [\mathbf{q}, g(\mathbf{q})]$ 
                SORT( $\mathbf{L}$ )
              }
          }
        }
      }
    }
  }
   $\mathbf{L} \leftarrow \mathbf{L} \setminus [\mathbf{p}, g(\mathbf{p})]$ 
   $e(\mathbf{p}) \leftarrow 1$ 
}

```

The point \mathbf{p} , being the seed point, with minimum global cost $g(\mathbf{p})$ is then removed from the sorted list \mathbf{L} and checked for unprocessed neighbors, which satisfy the condition $e(\mathbf{q}) = 0$. For each of the \mathbf{q} pixels the global cost function $g(\mathbf{q})$ is computed as the sum of the global cost $g(\mathbf{p})$ (where \mathbf{p} is a central point

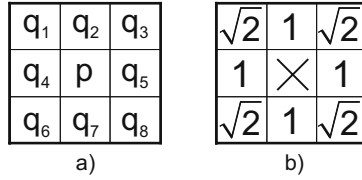


Fig. 1. Geometrical relations in neighborhood (a) pixel description, (b) geometrical distance

of the neighborhood) and the weighted local cost from \mathbf{p} to \mathbf{q} , hence $l(\mathbf{p}, \mathbf{q}) \cdot (1 \text{ or } \sqrt{2})$. The weight depends on the geometrical distance between \mathbf{p} and \mathbf{q} , according to Fig. 1. If a neighboring pixel \mathbf{q} along with its associated global cost $g(\mathbf{q})$ is in the sorted list L its previous global cost value $g(\mathbf{q})$ is compared with actually computed $g_{tmp}(\mathbf{q})$ and if the condition $g_{tmp}(\mathbf{q}) < g(\mathbf{q})$ is satisfied the image pixel is added to the sorted list L with a new cost value. Thus, if a pixel \mathbf{q} is not in the sorted list L , it is added to the list, with its global cost value $g(\mathbf{q})$ for latter processing. In both cases the array \mathbf{k} describing the tree is updated.

Finally, the pixel \mathbf{p} is removed from sorted list L , because each image pixel can be analysed only once. All analysed pixels are marked in the array \mathbf{e} (hence $e(\mathbf{p}) \leftarrow 1$). The algorithm runs until the sorted list L is not empty.

A path searching algorithm is similar to the commonly known Dijkstra algorithm [5]. The tree \mathbf{k} resulting from the Algorithm 2.1 constitutes a set of optimal paths, which connect each image pixel with the seed point. Therefore, the seed point can be achieved from any image pixel along an optimal path (path with the smallest cost). This kind of paths correspond to any image edges.

Once the path matrix is finished, a boundary path can be chosen dynamically via a free point. Interactive movement of the free point by cursor position causes the boundary to behave like a Live-Wire as it adapts to the new minimum cost path [6]. As soon as the advised by the user points are situated at the boundary of the segmented object, the displayed part of boundary should overlap with the actual segmented object boundary. Unfortunately, nature of the digital images, local noise, various shapes of segmented structures and wrong selection of the characteristic points, which does not consider the distance and curvature between these points, may result in delineating a random path away from the object boundary. In the traditional Live-Wire method only an interactive shifting of the free point towards the seed point can eliminate this weakness. This increases the user interaction requirement.

2.3 Live-Wire-on-the-Fly

Graph searching algorithm (Dijkstra algorithm [5]) is the most time consuming phase of the Live-Wire method. From the image segmentation point of view a full set of optimal paths including all image pixels is not necessary.

Algorithm 2.2. *Live-Wire-on-the-fly*(l, s, g, e, k, L, f)

% f – free point

Input: l, s, g, e, k, L, f

Data structures: $N(p)$

Output: g, e, k, L

Main:

if $e(s) = 0$

then $\begin{cases} g(s) \leftarrow 0 \\ L \leftarrow L \cup [s, g(s)] \end{cases}$

if $e(f) = 0$

then $\begin{cases} \text{while } L \neq \{\} \\ \quad \begin{cases} [p, g(p)] \leftarrow L[x_1, g(x_1)] \\ \text{if } p = f \\ \quad \text{then } break \\ \text{for each } q \in N(p) \\ \quad \quad \begin{cases} \text{if } e(q) = 0 \\ \quad \quad \quad \begin{cases} g_{tmp}(q) \leftarrow g(p) + l(q, x) \cdot \dots \\ \quad \quad \quad \quad \quad \dots \cdot (1 \text{ or } \sqrt{2}) \\ \text{if } q \in L \text{ and } g_{tmp}(q) < g(q) \\ \quad \quad \quad \quad \quad \text{then } \begin{cases} L \leftarrow L \setminus [q, g(q)] \\ g(q) \leftarrow g_{tmp}(q) \\ k(q) \leftarrow p \\ L \leftarrow L \cup [q, g(q)] \\ \text{SORT}(L) \end{cases} \\ \text{if } q \notin L \\ \quad \quad \quad \quad \quad \text{then } \begin{cases} g(q) \leftarrow g_{tmp}(q) \\ k(q) \leftarrow p \\ L \leftarrow L \cup [q, g(q)] \\ \text{SORT}(L) \end{cases} \end{cases} \\ \quad \quad \quad L \leftarrow L \setminus [p, g(p)] \\ \quad \quad \quad e(p) \leftarrow 1 \end{cases} \end{cases} \end{cases}$

The range of searched paths can be limited based on the user marked points: the seed and the free point.

Modification (Algorithm 2.2) relies on an additional stop condition, in the form of (*if $p = f$ then break*). If this condition is satisfied, the algorithm stops, the further analysis is possible and may be continue when the free point is moved to the area, where no path has been computed (i.e. $e(x) = 0$). This modification decreases significantly the computational complexity.

3 Modified Live-Wire Algorithm

In the traditional Live-Wire method the graph search algorithm (Dijkstra algorithm) has to process all image pixels. The computational complexity $\mathcal{O}(n^2)$,

where n denotes the number of graph vertices, is effected strongly by the size of the region subjected to the search algorithm. On the other hand, in most of the medical images the anatomical structures are surrounded by a background of no diagnostic importance.

It can be noticed, that homogeneous regions do not include object boundaries, which are delineated by the Live-Wire approach. A removal of these areas from the Live-Wire analysis reduces significantly the computational complexity. In our approach a fuzzy clustering method (FCM) has been implemented.

3.1 Wavelet Cost Map Definition

Local image contrasts are often more informative than light intensity values. A wavelet transform measures gray level image variations at different scales. Contours of image structures correspond to sharp contrasts and can be detected from the local maxima of the wavelet transform. Employing details from individual levels of wavelet transform cost map can be defined. It can be done in many ways. One approach was presented in [7, 8], the other is based on definition additional concepts. There are wavelet transform modulus defined by:

$$M_i = \sqrt{H_i^2 + V_i^2} \quad (3)$$

and wavelet transform angle defined by:

$$\varphi_i = \arg(H_i + j \cdot V_i) \quad (4)$$

where H_i , V_i are horizontal and vertical components from the i -level of wavelet decomposition, adequately.

In presented application only the components from the first level are respected and they are obtained with usage of Daubechies wavelets [9].

3.2 Fuzzy c-Means Clustering

Let $\mathbf{x}_k = (x_{k1}, \dots, x_{kn})$ be an observed data vector of $\{\mathbf{x}_k\}_{k=1}^N$ dataset in a feature space $\mathcal{F} \subset \mathbf{R}^n$. The standard FCM is derived to minimize the objective function [10]:

$$J(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^c \sum_{k=1}^N u_{ik}^m \|\mathbf{x}_k - \mathbf{v}_i\|^2 \quad \mathbf{x}_k, \mathbf{v}_i \in \mathcal{F} \quad (5)$$

with respect to the partition matrix element u_{ik} , the centre of the i -th cluster (\mathbf{v}_i) and for a given fuzzification level m ($1 < m < \infty$).

The partition matrix elements satisfy: $0 \leq u_{ik} \leq 1$, $\sum_{i=1}^c u_{ik} = 1 \forall k$, and $0 < \sum_{k=1}^N u_{ik} < N \forall i$.

The FCM clustering is performed interactively, starting with a set of c initially given prototypes and fuzzyfication level m . In each step a new partition matrix \mathbf{U} is created, satisfying:

$$u'_{ik} = \begin{cases} \frac{\|\mathbf{x}_k - \mathbf{v}_i\|^{-\frac{2}{m-1}}}{\sum_{z=1}^c \left(\|\mathbf{x}_k - \mathbf{v}_z\|^{-\frac{2}{m-1}} \right)}, & \text{if } I_k = 0 \\ \text{any, satisfying } \sum_{i \in I_k} u_{ik} = 1, & \text{if } i \in I_k \\ 0, & \text{otherwise} \end{cases}, \quad (6)$$

where $I_k = \{i : 1 \leq i \leq c \wedge \|\mathbf{x}_k - \mathbf{v}_i\|^2 = 0\} \forall 1 \leq k \leq N$.

The membership matrix \mathbf{U} is then employed to compute a new set of prototypes:

$$\mathbf{v}'_i = \frac{\sum_{k=1}^N u_{ik}^m \cdot \mathbf{x}_k}{\sum_{k=1}^N u_{ik}^m} \quad (7)$$

The procedure is repeated until the desired accuracy of \mathbf{V} is obtained, i.e. $\max(|\mathbf{V}' - \mathbf{V}|) < \epsilon$.

In many applications the objective function is modified yielding new approaches as FCM with spatial constrain [11] or geometrical guided FCM [12, 13].

3.3 Live-Wire Modification

Implementation of the FCM procedure requires an allocation of the number of classes. It is found experimentally and ranges from 7 to 10. Once an original slice (Fig.2a) is subjected to the FCM procedure, anatomical structure of similar intensity belong to one class (Fig.2b). It can be easily noticed, that homogeneous regions are clustered into one class and pixels around edges are assigned to various class. Only the latest are subjected to the search algorithm. Furthermore, only pixels linked with the seed point are examined by the searching conditions

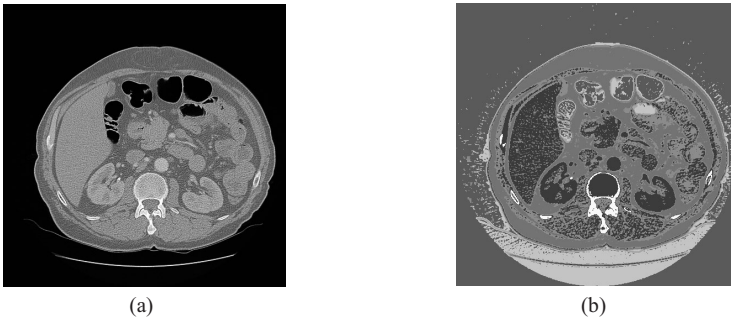


Fig. 2. Image classification (a) original image (b) graphical representation of classified image (FCM algorithm with $c = 7$)

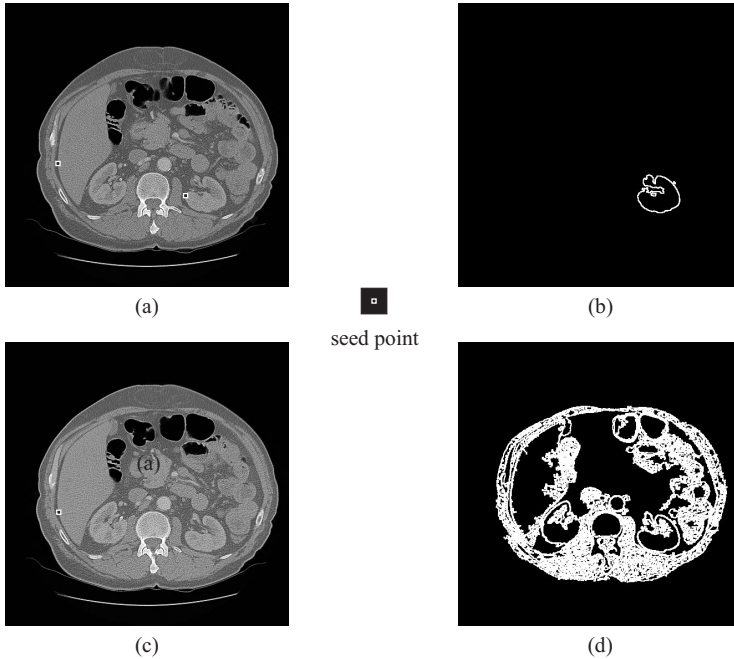


Fig. 3. Essential pixels of images (a), (c) original images with seed point (b), (d) essential pixels

(Fig. 3). When the anatomical structure to be segmented is well separated from other structures (kidney in Fig. 3a), the number of processed pixels (graph vertices) may be restricted to the neighbor pixels of the structure edge (Fig. 3b). However, if a structure overlaps with its neighbors structures (liver in Fig. 3c) the number of graph vertices increase, yet it is still significantly smaller in comparison with the entire slice size.

The new algorithm is shown in Algorithm 3.1 and called Live-Wire-FCM. The FCM preprocessing step allows a pixel \mathbf{p} to be selected for further processing only if its neighbor \mathbf{q} belongs to a different class. Thus, the condition **if** $e(\mathbf{q}) = 0 \dots$ in Algorithm 3.1 is replaced by **if** $e(\mathbf{q}) = 0$ and $I^{FCM}(\mathbf{q}) \neq I^{FCM}(\mathbf{p})$.

Pixels marked in black (Fig. 3b, d) will not be processed neither, when chosen as a seed point nor by the search procedure. The first case may be found as a disadvantage, yet the second one adds two new and important features to the method. Firstly, it reduces the computational complexity, secondly, it prevents shortcuts to be found, when a short path within homogeneous regions is at a lower cost than a long path along the edge.

The presented modification employs the on-the-fly mechanism. The graph searching algorithm is stopped, when the free point is reached. The method is presented in Algorithm, 4.1.

Algorithm 3.1. *Live-Wire-FCM*(l, s, \mathbf{I}^{FCM})

% \mathbf{I}^{FCM} – classified version of image obtained from *FCM* algorithm

Input: l, s, \mathbf{I}^{FCM}

Data structures: $g, e, L, N(p)$

Output: k

Main:

$g(s) \leftarrow 0$

$L \leftarrow L \cup [s, g(s)]$

while $L \neq \{\}$

do	{	do	{	then	{	$[p, g(p)] \leftarrow L[x_1, g(x_1)]$ for each $q \in N(p)$ if $e(q) = 0$ and $I^{FCM}(q) \neq I^{FCM}(p)$ $g_{tmp}(q) \leftarrow g(p) + l(q, x) \cdot (1 \text{ or } \sqrt{2})$ if $q \in L$ and $g_{tmp}(q) < g(q)$ then { $L \leftarrow L \setminus [q, g(q)]$ $g(q) \leftarrow g_{tmp}(q)$ $k(q) \leftarrow p$ $L \leftarrow L \cup [q, g(q)]$ SORT(L) } if $q \notin L$ then { $g(q) \leftarrow g_{tmp}(q)$ $k(q) \leftarrow p$ $L \leftarrow L \cup [q, g(q)]$ SORT(L) } $L \leftarrow L \setminus [p, g(p)]$ $e(p) \leftarrow 1$
-----------	---	-----------	---	-------------	---	---

4 Results

Due to the user's interaction, the algorithm performs well for various anatomical structures. In this study a segmentation of two anatomical structures (the kidney and the liver) from the CT abdomen series has been tested. Ten series have been subjected to the analysis. The segmentation has been performed with four methods. First, the traditional Live-Wire approach has been compared with its Live-Wire-FCM modification, (Tab. 1).

Since in the traditional Live-Wire method all optimal paths are searched, the entire image is subjected to the analysis, i.e. 100% of pixels are considered. The implementation of the FCM algorithm reduces the searched pixels to the boundary neighborhood. The average number of pixels subjected to the analysis has been shown together with the standard deviation.

```

Algorithm 4.1. LW-FCM-BK(l, s, g, e, k, L, f, IFCM)

Input: l, s, g, e, k, L, f
Data structures:  $\mathbf{N}(\mathbf{p})$ 
Output: g, e, k, L

Main:
if  $e(\mathbf{s}) = 0$ 
  then  $\begin{cases} g(\mathbf{s}) \leftarrow 0 \\ L \leftarrow L \cup \{\mathbf{s}, g(\mathbf{s})\} \end{cases}$ 
if  $e(\mathbf{f}) = 0$ 
  then  $\begin{cases} \text{while } L \neq \{\} \\ \quad \begin{cases} [\mathbf{p}, g(\mathbf{p})] \leftarrow L[x_1, g(x_1)] \\ \text{if } \mathbf{p} = \mathbf{f} \\ \quad \text{then } \textit{break} \\ \text{for each } \mathbf{q} \in \mathbf{N}(\mathbf{p}) \\ \quad \quad \begin{cases} \text{if } e(\mathbf{q}) = 0 \text{ and } \mathbf{q} \neq I^{FCM}(\mathbf{p}) \\ \quad \quad \begin{cases} g_{tmp}(\mathbf{q}) \leftarrow g(\mathbf{p}) + l(\mathbf{q}, \mathbf{x}) \cdot \dots \\ \quad \quad \quad \dots \cdot (1 \text{ or } \sqrt{2}) \\ \text{if } \mathbf{q} \in L \text{ and } g_{tmp}(\mathbf{q}) < g(\mathbf{q}) \\ \quad \quad \quad \text{then } \begin{cases} L \leftarrow L \setminus \{\mathbf{q}, g(\mathbf{q})\} \\ g(\mathbf{q}) \leftarrow g_{tmp}(\mathbf{q}) \\ k(\mathbf{q}) \leftarrow \mathbf{p} \\ L \leftarrow L \cup \{\mathbf{q}, g(\mathbf{q})\} \\ \text{SORT}(L) \end{cases} \\ \text{if } \mathbf{q} \notin L \\ \quad \quad \quad \text{then } \begin{cases} g(\mathbf{q}) \leftarrow g_{tmp}(\mathbf{q}) \\ k(\mathbf{q}) \leftarrow \mathbf{p} \\ L \leftarrow L \cup \{\mathbf{q}, g(\mathbf{q})\} \\ \text{SORT}(L) \end{cases} \end{cases} \\ L \leftarrow L \setminus \{\mathbf{p}, g(\mathbf{p})\} \\ e(\mathbf{p}) \leftarrow 1 \end{cases} \end{cases} \end{cases}$ 

```

Table 1. Number of essential image pixels in *LW* and *LW-FCM* methods

Structure	Number of pixels %		
	<i>LW</i>	<i>LW-FCM</i>	
	<i>m</i>	<i>σ</i>	
Kidney	100	0.05	0.02
Liver	100	3.01	0.41

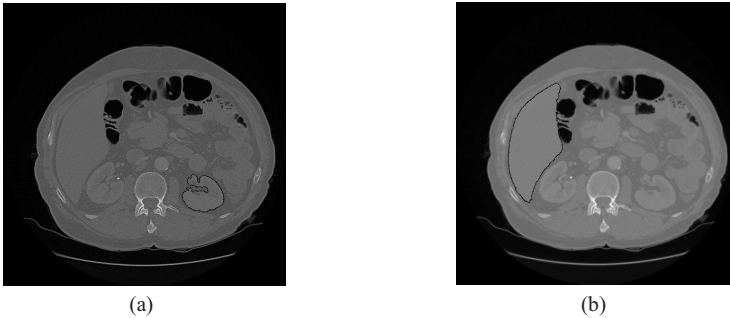
Implementation of the on-the-fly technique reduces the numerical complexity in the traditional Live-Wire as well as the Live-Wire-FCM, (Tab. 2). However, the results depend on the distance between the seed and the free points.

Table 2. Number of essential image pixels in *LW* on-the-fly and *LW-FCM* on-the-fly

Structure	Number of pixels %			
	<i>LW</i>		<i>LW-FCM</i>	
	<i>on-the-fly</i>	<i>on-the-fly</i>	<i>on-the-fly</i>	<i>on-the-fly</i>
	<i>m</i>	σ	<i>m</i>	σ
Kidney	17.52	4.76	0.03	0.01
Liver	28.63	5.47	1.93	0.27

Table 3. Number of characteristic points

Structure	Number of points	
	<i>LW</i>	<i>LW-FCM</i>
Kidney	6 – 9	3 – 5
Liver	8 – 13	5 – 9

**Fig. 4.** Segmentation results of (a) the kidney, (c) the liver

Additionally, the proposed modification shows certain advantage. Because pixels situated inside homogeneous regions are not included in the path searching procedure, thus paths crossing these areas are impossible. Thus, the total number of characteristic points, specified by the user is reduced, (Tab. 3).

5 Conclusions

This paper presents a modified version of the Live-Wire method employed to medical image segmentation. A Fuzzy C-Means procedure has been implemented in order to limit the performance of the analysis to the area surrounding the edges. The method has been evaluated on the basis of the CT abdomen images. The results have indicated a significant decrease of the computational complexity.

Moreover, a reduction of the area, subjected to the analysis has improved the segmentation accuracy, reducing at the same the number of points to be marked by the user.

References

1. Barrett, W.A., Mortensen, E.N.: Interactive Live-Wire Boundary Extraction. *Medical Image Analysis* 1(4), 331–341 (1997)
2. Mortensen, E.N.: Simultaneous Multi-Frame Subpixel Boundary Definition Using Toboggan-Based Intelligent Scissors for Image and Movie Editing. Brigham Young University (2000)
3. Mortensen, E.N., Barrett, W.A.: Intelligent Scissors for Image Composition. *Computer Graphics and Interactive Techniques*. In: ACM SIGGRAPH 1995, pp. 191–198 (1995)
4. Mortensen, E.N., Barrett, W.A.: Interactive Segmentation with Intelligent Scissors. *Graphical Models and Image Processing* 60(5), 349–384 (1998)
5. Dijkstra, E.W.: A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik* 1, 269–270 (1959)
6. Mortensen, E.N., Morse, B.S., Barrett, W.A., Udupa, J.K.: Adaptive Boundary Detection Using Live-Wire Two-Dimensional Dynamic Programming. In: *Computers in Cardiology*, Durham, pp. 635–638 (1992)
7. Więclawek, W.: Image segmentation – Live-Wire method. In: *International Workshop Control and Information Technology IWCIT 2005*, pp. 177–180 (2005)
8. Więclawek, W.: Live-Wire method with wavelet cost map definition for MRI images. In: *IFAC Workshop on Programmable Devices and Embedded Systems, PDeS 2006*, Brno, pp. 197–202 (2006)
9. Daubechies, I.: *Ten lectures on wavelets*. SIAM, Philadelphia (1992)
10. Bezdek, J.C.: *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers, Dordrecht (1981)
11. Pedrycz, W., Waletzky, J.: Fuzzy Clustering with Partial Supervision. *IEEE Transactions on Systems, Man, and Cybernetics, PartB* 27(5), 787–795 (1997)
12. Dulyakarn, P., Rangsanseri, Y.: Fuzzy C-Means Clustering Using Spatial Information with Application to Remote Sensing. In: *22nd Asian Conference on Remote Sensing, ACRS 2001*, vol. 1, pp. 212–215 (2001)
13. Kawa, J., Piętka, E.: Image Clustering with Median and Myriad Spatial Constraint Enhanced FCM, pp. 211–218 (2005)

The Clusterization Process in an Adaptive Method of Image Segmentation

Aleksander Lamza¹ and Zygmunt Wrobel²

¹ Department of Biomedical Computer Systems, Institute of Computer Science,
Faculty of Computer and Materials Science, University of Silesia in Katowice,
ul. Bedzinska 39, 41-200 Sosnowiec
aleksander.lamza@us.edu.pl

² zygmun.wrobel@us.edu.pl

Summary. In the article a new method of automatic image segmentation is presented. The aim was to eliminate the necessity of defining the number of outcome areas. Homogeneous areas take part in the growth process. The areas merge when the homogeneity condition is fulfilled. The threshold value changes during the segmentation process, fitting the changeable conditions.

1 Introduction

The image segmentation is a process of distinguishing areas of specific parameters in a given image [1, 2, 12, 13]. The method presented in this article is a region merging method, which is frequently used and modified [5, 7, 9, 10, 11]. We present a new automatic and adaptive method of generating merging areas, which eliminates the necessity of setting the number of outcome areas, defining the seeds and segmentation parameters. The growth occurs in homogeneous areas – continuous groups of pixels. During the algorithm's work the number of homogeneous areas decreases when areas merge with their neighbours. The algorithm stops when no more mergings can occur.

2 The Idea of Homogeneous Areas

Before presenting the process of automatic generating and merging of areas, we should define the concept of a homogeneous area.

A homogeneous area is a continuous group of pixels, which fulfill the similarity condition. The parameter describing similarity is calculated based on the variance of the gray level of the pixels [6, 10], which is described in details further. A group of pixels building a homogeneous area is labeled as H_m , where: m – the number of the following area. Each area is a set of pixels $H_m = \{\mu_1, \mu_2, \dots, \mu_i\}$, where μ_i – the gray level of a given pixel in the area m , i – the number of pixels in that area. According to that notation, we shall denote the gray level of a pixel i in a area m by $\mu_i(H_m)$.

As it was mentioned before, the pixels building a area should be continuous, which means that each element of a area should have at least one neighbour of the same area. The neighbourhood is understood like this (fig. 1):

-	μ_{n2}	-
μ_{n3}	μ_i	μ_{n1}
-	μ_{n4}	-

Fig. 1. The neighbourhood of area elements

For the element μ_i the neighbours are the elements $\mu_{n1}, \mu_{n2}, \mu_{n3}, \mu_{n4}$; while the elements "-" are not treated as neighbours. These are the conditions for areas:

- No area can be an empty set: $H_m \neq \emptyset$, where m is the area number, M – the number of areas.
- The sum of all areas must cover the whole image: $\bigcup_{m=1}^M H_m = X$, where m is the area number, M – the number of areas, X – the set of all elements (pixels) of an image.
- The areas can not have common elements: $H_m \cap H_n = \emptyset$, where m and n are area numbers, M – the number of areas.

3 The Algorithm

There are two phases in a proposed segmentation algorithm:

1. Initial processing – generating the initial set of areas.
2. Merging the areas – checking the homogeneousness condition and merging the areas, which have fulfilled that condition.

In the first phase the initial number of homogeneous areas is generated automatically. It's usually much bigger than expected, because some homogeneous areas sharing the same parameters, but in different parts of an image, can occur. The second step is the adaptive growth of individual areas, resulting from the merging of the areas that fulfill the homogeneousness condition. The number of areas decreases all the time, when successive areas merge with their neighbours. The algorithm stops when the remaining areas can not be merged.

3.1 Generating the Initial Set of Areas

At first the initial set of areas must be determined. Therefore the image is searched for continuous areas of pixels with similar gray levels. The image is scanned row by row, beginning with the pixel (0, 0). Such algorithm checking of the pixels' neighbours and merging them to the appropriate areas.

3.2 Merging the Areas

The main part of the proposed algorithm is the adaptive merging of the areas that fulfill the homogeneousness condition. The result is getting minimal number of areas for the given image. The algorithm works with iterations. Each iteration consists of following steps:

1. Random choosing area H_m out of the area set H ,
2. Checking the similarity condition between the area H_m and all its neighbours H_n ,
3. If areas H_m and H_n fulfill the homogeneousness condition, all elements of the neighbouring area $\mu(H_n)$ are merged to area H_m .

In the first step the algorithm randomly chooses an area, which will be later checked for similarity condition. Each element of an area stores information about its neighbourhood. After random choosing an area H_m a neighbourhood table for that area is created. To check the homogeneousness condition for every area H_n neighbouring the H_m , the variance of the sum of elements for both areas is calculated. The homogeneousness condition is estimated based on the gray levels variance of all elements for merged areas. The general formula for the variance of a single area is shown below (1):

$$\sigma^2(H_m) = \frac{1}{|H_m|} \sum_{i=1}^{|H_m|} \left[\mu_i(H_m) - \overline{\mu(H_m)} \right]^2, \tag{1}$$

where:

- $\sigma^2(H_m)$ – the variance of the elements in the area H_m ,
- $|H_m|$ – the number of the elements in the area H_m ,
- $\mu_i(H_m)$ – the gray level of the element i from the area H_m ,
- $\overline{\mu(H_m)}$ – the average gray level of the elements in the area H_m .

In a proposed algorithm the variance of the two areas' sum (the chosen one's and its neighbour's) is always determined.

$$\sigma^2(H_m, H_n) = \sigma^2(H_m \cup H_n) \tag{2}$$

The variance is calculated from the equation (3, 4):

$$\sigma^2(H_m, H_n) = \frac{1}{|H_m|+|H_n|} \left[\sum_{i=1}^{|H_m|} \left[\mu_i(H_m) - \overline{\mu(H_m, H_n)} \right]^2 + \sum_{j=1}^{|H_n|} \left[\mu_j(H_n) - \overline{\mu(H_m, H_n)} \right]^2 \right], \tag{3}$$

$$\overline{\mu(H_m, H_n)} = \frac{1}{|H_m| + |H_n|} \left[\sum_{i=1}^{|H_m|} \mu_i(H_m) + \sum_{j=1}^{|H_n|} \mu_j(H_n) \right], \tag{4}$$

where:

- $\sigma^2(H_m, H_n)$ – the variance of the sum of the elements in H_m and H_n ,
- $|H_m|, |H_n|$ – the number of the elements in the area H_m and H_n ,

$\mu_i(H_m), \mu_i(H_n)$ – the gray level of the element i (j) from the H_m (H_n),
 $\overline{\mu(H_m, H_n)}$ – the average gray level of all elements in the areas H_m and H_n .

When the variance of the area H_m and all its neighbours H_n is calculated, the algorithm chooses a neighbouring area for which the lowest value of the variance was estimated. Then, it checks the global condition, determining if the estimated variance is smaller than the threshold value σ_{max}^2 :

$$\sigma^2(H_m, H_n) < \sigma_{max}^2 \quad (5)$$

If the condition is filled, the merging process begins. All elements of the area H_n : $\mu(H_n)$ are transferred to the area H_m , and then the area H_n is removed from the area set H . If the condition is not filled, the areas are not merged and the algorithms jumps to the first step: the random choosing of another area.

3.3 The Adaptation of the Threshold Variance

In the presented algorithm the parameter responsible for determining if the areas will merge or not is the threshold variance σ_{max}^2 . Up to this point it is been treated as constant. In that case for every image, regardless of objects size and texture, the segmentation would proceed in a fixed way and the results would not be satisfactory. A possible solution would be to estimate the threshold variance based on the image content, but the main purpose was to avoid any external parameters. To automatize the process a model of the threshold variance adaptation based on the segmentation process dynamics was developed.

As mentioned before, the value of the variance σ_{max}^2 has to change during the segmentation process. In a case when, in a given iteration, an area can not merge with any of its neighbours, the threshold value should be adequately increased to allow the merging. On the other hand, immediate increasing this value to the needed level would result in merging all the areas in a one area covering the whole image. This can not happen.

Therefore, a model of variance adaptation should change the value σ_{max}^2 smoothly, not allowing it to increase excessively. Thus – when the merging does not occur (the homogeneousness condition is not fulfilled), the threshold variance increases according to equation (6):

$$\sigma_{max}^2(i) = \sigma_{max}^2(i-1) \cdot \left[1 - \exp\left(-\frac{i}{\tau_1}\right) \right], \quad (6)$$

where:

i – the number of the current iteration,

τ_1 – the time constant of the variance increase (expressed in iterations).

If the areas merge in the current iteration, the value of threshold variance decreases until the next unsuccessful merging occurs.

$$\sigma_{max}^2(i) = \sigma_{max}^2(i') \cdot \left[1 - \exp\left(-\frac{i-i'}{\tau_2}\right) \right], \quad (7)$$

where:

- i – the number of the current iteration,
- i' – the iteration, for which the last successful merging occurred,
- τ_2 – the time constant of the variance decrease.

The value of the variance σ_{max}^2 is calculated in every step of the algorithm. The time constants τ_1 (increase) and τ_2 (decrease) have their values determined experimentally. The variance σ_{max}^2 oscillates during the areas' merging. At the end of the segmentation process a rapid fluctuation is visible. This sudden increase of the threshold value results in merging the areas which should not be merged. Therefore, the large area of an image is "flooded" by a single area. This increase was triggered by small number of areas and a high value of variance. The unsuccessful mergings occur often, resulting in a constant increase of the threshold variance value. To avoid such situations, a parameter correcting the time constant of the variance increase (τ_1) was introduced. During the segmentation process the number of areas decreases successively. Along with that decrease, the time constant τ_1 should increase to induce a slower rate of the threshold variance growth.

$$\tau_1 = \tau_0 \cdot M_0 \cdot \frac{1}{M_i} \quad (8)$$

where:

- τ_1 – the time constant of the variance increase,
- τ_0 – a parameter determined experimentally,
- M_i – the number of the areas in the iteration i ,
- M_0 – the initial number of the areas.

After introducing the correcting parameter much better results were achieved. Regardless of the input image characteristics, three phases can be distinguished in the course of the variance during successive iterations. Figure 2 illustrates this fact. Three values of the variance are marked on the graphs: σ_p^2 – the variance for the merged areas, σ_n^2 – the variance for the unsuccessful merging and σ_{max}^2 – the threshold variance. The phase I – merging (fig. 2-a) is characterized by a considerable frequency of changes in the merged areas variance σ_p^2 with rare fluctuations of the variance σ_n^2 . It should be mentioned, that when the growth of the variance σ_n^2 occurs, the value of σ_{max}^2 equals the variance of unmerged areas after a few iterations. In the phase II – stabilization (fig. 2-b) – areas merge sporadically. The variance σ_n^2 fluctuations occur much more often, yet they are rather homogeneous. The value of the variance σ_{max}^2 does not reach the maximum because of the increase of the time constant τ_1 . The phase III – separation (fig. 2-c) is characterized by a visible distinction between the variances σ_n^2 and σ_{max}^2 . The value of the variance σ_p^2 is at its lowest level and at some point it disappears, which means the process of the areas merging is done.

The merging of big areas is inadvisable, because in such cases inaccuracies in defining segments may occur. It does not happen in the first and the second phase, because the threshold variance value is then relatively low. Yet, in the

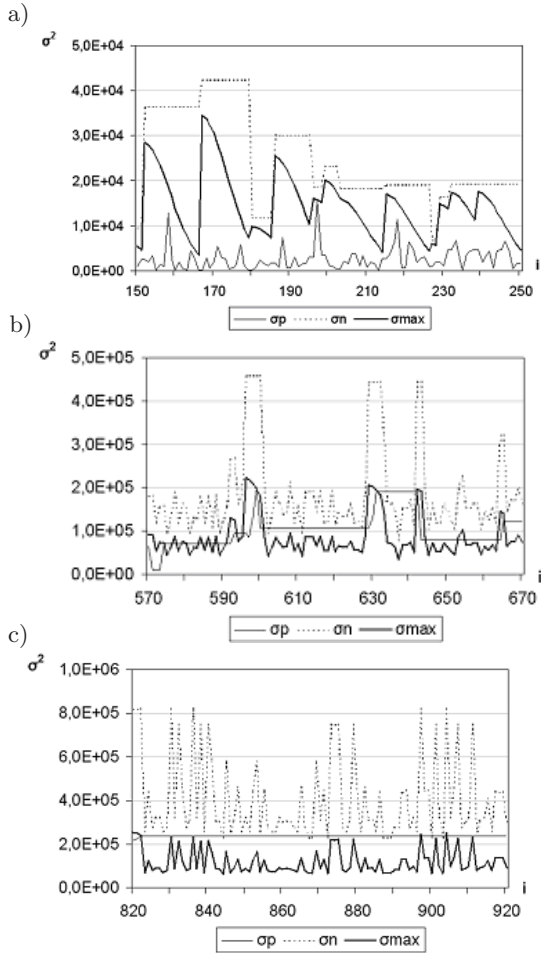


Fig. 2. The values of variance in successive iterations for three phases of segmentation

third phase the merging of big areas may at some point happen, despite of the growth of the variance increase time constant. Therefore, a modification of the areas choosing algorithm in the third stage was proposed. The areas are no more chosen by chance: only the biggest ones are checked for their possibility to merge with the smaller, neighbouring ones. This stopped big areas from merging and precipitated the end of the segmentation process.

3.4 The Edge Strength Factor

To avoid situations when areas divided by an edge get merged, the edge strength factor was introduced. The input image is filtered by an edge detection filter. Various types of masks were tested and it came out that the most effective one is the Sobel mask [3, 12]. A set of two masks for two main directions was applied.

The resulting image can be treated as a table of edge strength factors. After rescaling, the values of edge strength for individual pixels are between 0 and 1, where 0 means no edge, and 1 – the most distinct edge. For the purpose of the method's description, the edge strength of two neighbouring areas can be represented as a set: $E(H_m, H_n) = \{e_1, e_2, \dots, e_l\}$, where l is the number of bordering elements between the areas H_m and H_n . The edge strength is taken into account during the calculation of two areas' variation. The equation (3) should be therefore modified:

$$\sigma^2(H_m, H_n) := b \cdot \sigma^2(H_m, H_n), \quad (9)$$

where:

$b = \sum_{e \in E(H_m, H_n)} e$ – the edge strength factor,
 e – the values of edge strengths for bordering elements of H_m and H_n .

4 Conclusions

An important problem when analyzing real images is the detail level of segmentation [4, 8]. The example of segmentation microscopy image is shown below (fig. 3). The segmentation of the whole image resulted in distinguishing three areas (the background and two nucleuses). The segmentation of the nucleus resulted in a much bigger amount of details.

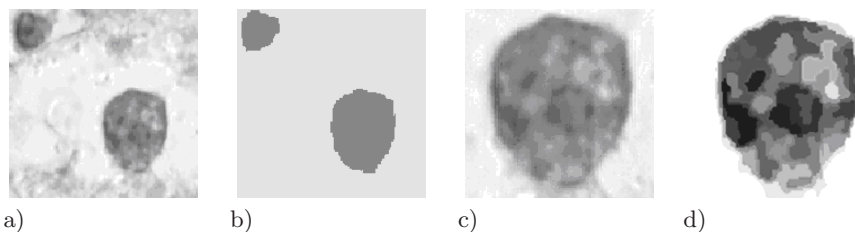


Fig. 3. Microscopy image (a, c) and segmented image (b, d)

The presented segmentation method were tested on various classes of images and the results are satisfactory. The detail level of the segmentation depends on the image size, therefore a general segmentation can be completed with a detailed one, proceeded on chosen, earlier isolated parts of an image.

References

1. Cour, T., Benezit, F., Shi, J.: Spectral segmentation with multiscale graph decomposition. *Computer Vision and Pattern Recognition* 2, 1124–1131 (2005)
2. Dhawan, A.: *Medical Image Analysis*. Wiley-IEEE Press (2003)
3. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*. Prentice Hall, New Jersey (2002)

4. Hoover, A., Jean-Baptiste, G., Jiang, X.: An Experimental Comparison of Range Image Segmentation Algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence* 18(7), 673–689 (1996)
5. Lamecker, H., Lange, T., Seebß, M.: Automatic Segmentation of the Liver for Preoperative Planning of Resections. *Proceedings MMVR, Studies in Health Technologies and Informatics* 94, 171–174 (2003)
6. Lamza, A., Trela, M., Bialy, S., Wrobel, Z.: Agent-based method of automatic image segmentation with dynamic threshold. *IEEE Signal Processing*, 71–74 (2004)
7. Pantofaru, C., Hebert, M.: A Comparison of Image Segmentation Algorithms. Technical Report CMU-RI-TR-05-40, Robotics Institute, Carnegie Mellon University (2005)
8. Porikli, F.: Automatic Image Segmentation by Wave Propagation. In: *SPIE Conference on Image Processing: Algorithms and Systems III*, vol. 5298, pp. 536–543 (2004)
9. Soille, P.: *Morphological Image Analysis: Principles and Applications*. Springer, Heidelberg (2004)
10. Veenman, C.J., Reinders, M.J.T., Backer, E.: A Cellular Coevolutionary Algorithm for Image Segmentation. *IEEE Transactions on Image Processing* 12(3), 304–316 (2003)
11. Wang, H., Suter, D.: A Model-Based Range Image Segmentation Algorithm Using a Novel Robust Estimator. In: *3rd International Workshop on Statistical and Computational Theories of Vision*, Nice, France (2003)
12. Wrobel, Z., Koprowski, R.: *Praktyka przetwarzania obrazu w programie MATLAB*. Wydawnictwo Exit, Warszawa (2004)
13. Zielinski, K.W., Strzelecki, M.: *Komputerowa analiza obrazu biomedycznego. Wstep do morfometrii i patologii ilosciowej*. Wydawnictwo Naukowe PWN, Warszawa – Lodz (2002)

Content-Based Indexing of Medical Images for Digital Radiology Applications

Piotr Boninski and Artur Przelaskowski

Warsaw University of Technology
pboninsk@ire.pw.edu.pl

Summary. This paper concerns content-based image retrieval in medical domain considering the challenges of rapidly growing amounts of medical data, permanent progress of computer-aided radiology and the development of global data exchange networks (like Mammogrid). The aim of featured research was to propose effective content-based image retrieval algorithms in two approaches to that problem. The first is medical images indexing for various modalities, on the assumption that we do not attempt to analyze image semantics. In that approach we try to find the images 'visually similar' to the given one – with similar organ, modality, orientation, etc. The second approach undertaken in conducted research is medical images indexing with taking into consideration their semantics. Such approach makes use of 'domain knowledge' about specified modality, examination, giving the opportunity to introduce descriptors of image semantics, especially related to diagnostic content. The methods and algorithms characterized in this paper are both related to various modalities and strictly dedicated to the one modality only - in this research it is mammography. The obtained results show the usefulness of proposed methods.

1 Introduction

Medical images contain vital information about patient state. This information can be used to make diagnosis and to facilitate therapeutic and surgical treatments. Traditionally, these images are stored on films. During the past couple of years, development of digital technologies dramatically changed the approach. The typical hospital is able to produce gigabytes of image information per day and terabytes per year. Effective management of these huge image databases requires archiving and communication system (PACS) [4].

Although a PACS relies on complex, probably a bit overwhelmed protocols such as DICOM, image selection within a DICOM network is based currently on alphanumerical information only. However, information contained in medical images is much more complex than that residing in alphanumerical format, hence the information provided by DICOM structures is not enough to find image data efficiently [2]. The problems with textual description of showed and interpreted information lead to the idea of using content-based indexing and

retrieval methods that might have a great influence on effectiveness of PACS systems. By means of uniform protocols and the API provided by the PACS core, this integration can be realized to satisfy the right-information-paradigm maintaining the autonomy of both components, the PACS and content-based image retrieval (CBIR) system [3].

The other problem related to modern PACS and image databases in general, is efficient communication access to large amount of data. The JPEG2000 standard describes effective tools for progressive and interactive image data transmission in medical imaging applications: PACS-RIS-HIS enterprise, teleradiology and CAD utilities. However, optimization of wavelet data transformation, selection and stream forming procedures can significantly improve standard implementations available nowadays in the market. Diagnostic quality enhancement and accelerating coding process of applied compression tools can actually improve image-oriented real-time diagnostic systems [5]. Original contribution of our work is an IShark, which is experimental PACS system, with additional support of content-based image retrieval system, distributed search and JPEG 2000-based image communication with interactive protocols. Such approach gives very flexible and effective image retrieval tool.

This paper presents two approaches to content-based indexing. The first may be considered as classical one. We try to build an index which is not aware of an image semantic, it just try to cover as many as possible of image properties – like color, texture, shape, etc. Such approach could be considered as 'classical' one, as it is hard to introduce a semantic-driven descriptor, especially for wider range of images.

The second approach presented in the paper is a try to propose a semantic-aware descriptors, which make use of domain knowledge and gives an opportunity to have a real content-aware indexing technique for particular image type.

2 Materials and Methods

The presented algorithms and ideas was implemented in prototype IShark system.

2.1 IShark System

The IShark consists of 4 main elements:

1. database environment with user interface
2. content-based image retrieval system
3. JPEG 2000 interactive codec
4. web service for distributed searching

Additionally, there is DICOM data importer, which is able to transfer data between DICOM-compliant devices, like separate imaging systems (e.g. tomographs, mammographs) or integrated PACS.

Database Environment with User Interface

The main element is database environment networked to DICOM image source (i.e. PACS or imaging systems). It offers most features expected from such systems. The data are organized into series/study/patient tree, so, from that point of view, it could be treated as simple PACS. Generally, IShark system is a reference database system, which is a good support for classic PACS environment. If physician want to find a case similar to the selected one – IShark provides an efficient ways to do so. The user has generally two ways to find an interesting case. The first one – using classic text-based approach. The novel element in IShark is that there is possibility to make distributed queries. Every IShark database has dedicated webservice, which is registered at central web service. A second possibility is to search the database for diagnostically similar cases using content-based image retrieval engine.

Database environment is also a place where user logs in. It has web-based user interface with heavy use of asynchronous requests (known as AJAX) to make the interface convenient for the end user.

JPEG 2000

Image transmission between database module and the user (diagnostic workstation, tele-radiologic application) is made by Java applet, which is a front-end to a novel JPEG 2000 codec implementation. More details about this codec and its useful properties can be found in [6].

IShark Webservice

A standalone IShark database module should be "attached" to webservice, which provides search functionality for main module. All search requests from PACS go to the webservice, which makes query to a database. Such architecture gives an opportunity to search other ISharks, for example located at other medical centers.

When the physician wants to make "distributed" query, PACS module sends it to central webservice, which has a list of all IShark services. The query is then resend to all that services, collected and returned to the client. Such approach gives a possibility to create distributed image databases, encompassing many autonomic databases.

2.2 Features for Retrieval of Various Modalities Images

Image-Domain Features

1. Gray-level histogram. We found classical gray-level histogram as useful and easy to compute feature for content-based retrieval. A problem with histograms is the discontinuity. That is, slightly changing the image might change the bin assignments and thus the resulting histogram completely. To overcome this problem fuzzy histograms can be used. The goal of fuzzy

histograms is to remove the discontinuous bin assignment of the traditional histogram. We evaluated both traditional and fuzzy histograms.

2. Tamura features. In [7] the authors propose texture features corresponding to human visual perception. We found a subset of originally proposed features as useful for medical images. These features are: coarseness, contrast, and directionality.
3. MPEG-7 features. MPEG-7 Visual Standard specifies a set of descriptors that can be used to measure similarity of images or video. The significant role in MPEG-7 descriptors play a texture-based descriptors. We evaluated Edge Histogram, Homogeneous Texture and Texture Browsing Descriptor. The details about these are presented in [1].

Wavelet-Domain Features

There is a common problem how to find a right region to evaluate local features. The most often used segmentation method is dividing image into parts (blocks) with fixed size range, e.g. quad-tree segmentation or contour-based segmentation. Another methods relies on a set of detected single points or a small group of points, e.g. corners, around which the interesting part of image is assumed to be.

The detailed algorithm with all details is presented in [1], here we provide only brief description.

- We start with create a wavelet image representation. Then take the coarsest part of multiscale hierarchy (lowest frequency),
- For each wavelet coefficient, find the maximum n child coefficients,
- Track it recursively in finer resolutions,
- At the finest resolution, set the saliency value of the tracked pixel: the previous value and the sum of the wavelet coefficients tracked,
- Threshold to extract the most prominent points.

Such approach gives us an opportunity to bring out the points with most significant value for the image. We designed a descriptor basing on proposed semantic points. We call it Matched Salient Regions as it basically represents a total distance of matched regions between two images. The region means a neighborhood around found semantic point. More details about the algorithm can be found in [1].

2.3 Features for Retrieval of Mammographic Image

Beside the proposed general-purpose image features, we propose also a content-aware approach. As such approach intensively makes use of domain knowledge, it has to be limited to particular image type. We decided to develop the descriptors for mammographic images as it is very important and hard to analyze modality. The content of medical image is strictly related to diagnostically important structures (morphological or functional), lesions, pathology symptoms etc.

There are two main pathology symptoms in mammography [1]: microcalcifications and masses. We proposed the descriptors for both of them, additionally we evaluated and indexed breast density.

Breast density descriptor is just a one scalar data, which is breast density in BIRADS scale (from I – fat tissue to IV – dense tissue).

Microcalcification Descriptor

The microcalcification detection algorithm, used to compute proposed descriptor, is out of scope of this paper and could be found in our other papers [1]. The main points of the descriptor extraction algorithm are as follows:

- Preliminary processing
- Convolutions with Laplacian filters of different scales for localization of bright spots
- Potential microcalcification segmentation
- Clustering with DBSCAN-based algorithm
- Feature extraction of microcalcification clusters
- Descriptor evaluation

The descriptor of microcalcification cluster is as follows:

1. Cluster localization
2. Size
3. Brightness
4. Cluster shape
5. Layout regularity of microcalcifications in cluster
6. Microcalcification edge irregularity

Mass Descriptor

Similar to the previous section, we do not discuss all details of mass detection algorithm, but we present only the descriptor of the potential mass. The main points of descriptor extraction algorithm are as follows:

1. Breast segmentation
2. Rayleigh transform function to emphasizing candidates for mass area
3. Region-growing based segmentation to recover approximate mass shape
4. Descriptor evaluation

The elements of mass descriptor are as follows:

1. Mass localization
2. Size
3. Brightness
4. Mass shape
5. Mass edge irregularity

3 Testing Scenarios and Results

3.1 Various Modalities Images Indexing

During our research we found a few descriptors set as the most effective. The proposed description sets and the weights used for every described feature were given in Table 1. The precision/recall graphs for proposed descriptors were presented in Fig. 1.

Table 1. Feature sets for proposed descriptors

Set number	Feature	Weight
I	Tamura textual features	3
	Edge Histogram	6
	Matched salient regions	2
II	Tamura textual features	3
	Edge Histogram	6
III	Global textual descriptor	2
	Homogeneous Texture Descriptor	4
	Matched salient regions	2
IV	Tamura textual features	2
	Fuzzy gray-level histogram	4
	Matched salient regions	2
V	Tamura textual features	2
	Gray-level histogram	4
	Matched salient regions	2

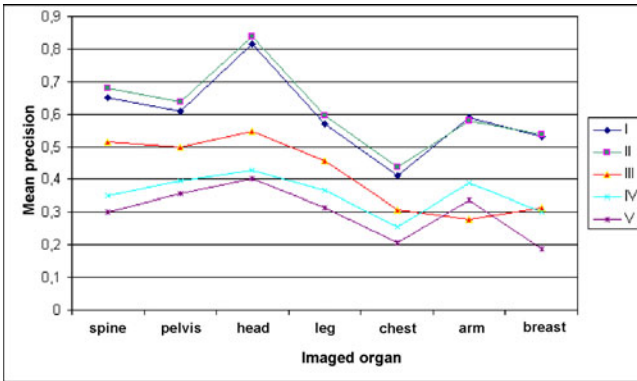


Fig. 1. Average retrieval precision for selected body regions for proposed descriptor sets

3.2 Mammographic Images Indexing

Evaluating of mammographic images retrieval is more complicated than retrieval of various modalities. We do not have here a clean criteria to decide whether the

answer is relevant or not. This fact implies that one should define relevancy criteria. We propose following relevancy criteria for mammographic images:

1. Mass - at least one mass in relevant images
2. Malignant mass - at least one malignant mass in relevant images
3. Two masses - at least two masses in relevant images
4. Spicular mass - at least one spicular mass in relevant images
5. Microcalcifications cluster - at least one microcalcifications cluster in relevant images
6. Malignant microcalcifications cluster - one malignant microcalcifications cluster in relevant images
7. Pathology symptoms in thin breast (BIRADS I II) - BIRADS I or II density with at least mass or microcalcifications cluster in relevant images
8. Pathology symptoms in dense breast (BIRADS III IV) - BIRADS III or IV density with at least mass or microcalcifications cluster in relevant images

These criteria were formulated to mimic real scenarios in clinical practice. The precision/recall graph was presented in Fig. 2.

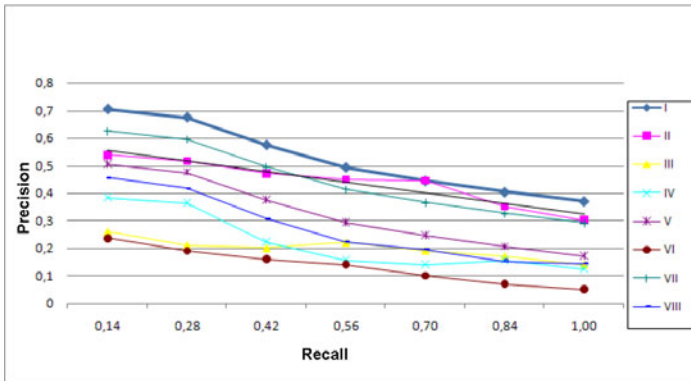


Fig. 2. Retrieval efficiency for proposed scenarios

4 Conclusions

The presented system consists of several modules that provide far more efficient way of interaction with image database than classic text-based queries from PACS system. The important thing is that IShark does not take out any search functionality known from classical systems. The new features concentrate mainly on distributed search, content-based search and effective, interactive JPEG 2000 image communication. The distributed searching feature could be very helpful for a user. It gave easy possibility of sharing interesting cases between medical centers, e.g. for better diagnosis or educational purposes. Of course the sent data are anonymized and encrypted.

The content-based image retrieval is relatively new technique, especially in medical domain. The research related to CBIR is still ongoing, not only in our group, so the results presented here are definitely preliminary. The future research will be concentrate on wavelet-domain features for more pathology-oriented retrieval of selected modalities, other that mammography.

References

1. Boninski, P.: Indeksowanie obrazow medycznych na potrzeby radiologii cyfrowej, rozprawa doktorska, Politechnika Warszawska (in Polish) (2007)
2. Güld, M.O., Kohnen, M., Keysers, D., Schubert, H., Wein, B., Bredno, J., Lehmann, T.M.: Quality of DICOM header Information for Image Categorization. In: SPIE Intl. Symposium on Medical Imaging, San Diego, CA. Proceedings SPIE, vol. 4685, pp. 280–287 (2002)
3. Lehmann, T.M., Wein, B.B., Greenspan, H.: Integration of Content-based Image Retrieval to Picture Archiving and Communication Systems, MIE 2003 Proceedings (2003)
4. Müller, H., Michoux, N., Bandon, D., Geissbuhle, A.: A review of content-based image retrieval systems in medical applications—clinical benefits and future directions, r. *International Journal of Medical Informatics* 73, 1–23 (2004)
5. Przelaskowski, A., Hałasa, P., Rieves, D.D.: Progressive and interactive modes of image transmission: optimized wavelet-based image representation. In: 3rd International Conference on Telemedicine and Multimedia Communication, Abstract Book, pp. 65–66. Kajetany (2005)
6. Schaefer, G.: JPEG vs. JPEG2000 from an image retrieval point of view. In: Proc. IEEE Int. Conference on Image Processing, Singapore, pp. 437–440 (2004)
7. Tamura, H., Mori, S., Yamawaki, T.: Texture features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics* 8, 460–473 (1978)

Shape and Texture Feature Extraction for Retrieval Mammogram in Databases

Ryszard S. Choraś

Faculty of Telecommunications
University of Technology & Life Sciences
85-796 Bydgoszcz, S. Kaliskiego 7
Poland
choras@utp.edu.pl

Summary. The huge amount of digital images generated in hospitals leads to the need of automatic storage and retrieval of them. A *Picture Archiving and Communication System* (PACS) should incorporate properties allowing to retrieve these images and adding *Content-Based Image Retrieval* (CBIR) capabilities to PACS makes it more powerful to assist diagnosis. Such systems provide features which combine color, shape and spatial features to query an image. In response to a user's query, the system returns images that are similar in some user-defined sense. Our purpose in this study is to develop a method of image mammogram feature extraction (microcalcifications and masses features) in CBIR system.

1 Introduction

Now medical imaging systems produce more and more digitized images in all medical fields. These images are interesting for diagnostics but access to huge database requires efficient indexing to enable fast access to images in databases.

In the picture archiving and communication systems (PACS) used in modern hospitals, the current practice is to perform automatic image indexing and to retrieve images based on image digital content (CBIR). CBIR methods employ image processing algorithms to extract relevant features from the images, organizing them as feature vectors. Then, indexed feature vectors can lead to fast and efficient image retrieval.

Recent years have resulted in the introduction of many domain-specific CBIR solutions for a large array of medical imaging modalities, such as mammography studies [1]. In this work we investigate the use of content-based image retrieval (CBIR) for digital mammograms. The goal of CBIR is to obtain from a possibly very large image database those images that are similar in content to an image of interest (the query image). A popular approach is to employ measured image features to provide a description of the content of the image (Fig. 1). As it is shown in Figure 1, the CBIR system architecture can be divided into two main blocks: an on-line block and an off-line block. In the on-line block the radiologist is studying a new mammogram. Next, in a fast way, the expert wants/needs to

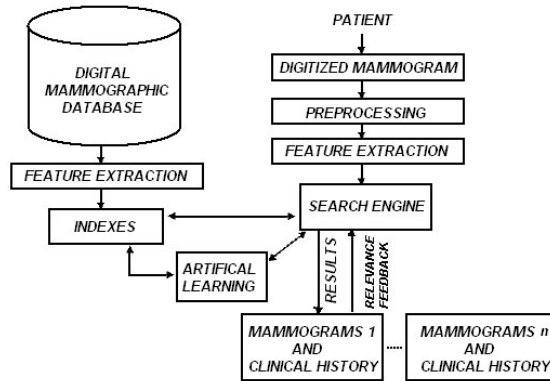


Fig. 1. Typical architecture of a CBIR system that retrieves similar diagnosed cases compared to an unknown one

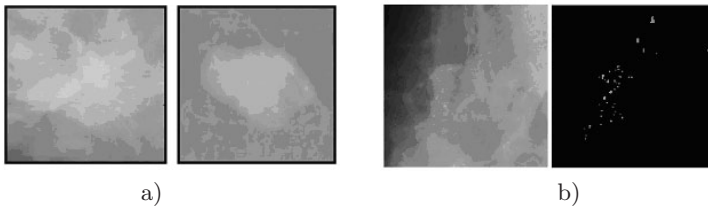


Fig. 2. Example of irregular mass (left side - malignant, right side - benign) a) and microcalcifications b)

find similar analyzed cases to this one. These similar cases are the basis of the off-line block, because they had been previously characterized, and stored into a huge database of cases.

Breast cancer are normally associated with:

- Asymmetry between images of left and right breasts.
- Distortion of the normal architecture of the breast tissue.
- Presence of masses in the breast.
- Presence of microcalcifications in the breast.

In this work we investigate a two types of objects that appear in mammogram images: masses and microcalcifications [3, 5]. A mass is a space-occupying lesion. There are several types of masses found in mammogram images [6]. Masses are categorized by their shape and density. Figure 2a illustrates some mass shapes found in mammogram images. Other important objects observed in mammograms are microcalcifications. These represent calcium deposits located in the breast tissue and are considered highly indicative of breast cancer. Microcalcifications appear as small, bright objects that stand out from the surrounding tissue [2, 7]. Figure 2b shows examples of microcalcifications.

2 Mammograms Feature Extraction

Finding the best features and getting the high classification rate is the main problem in classification of mammograms. The feature extraction consists of:

- ROI marked in mammograms,
- Feature extraction from ROI,
- Feature selection for the classification.

Mass detection in mammography is based on shape and texture based features [4]. Ten shape and texture based features were extracted from the masses after segmentation. The twelve features are listed below:

1. Mass area. The mass area, $A = |R|$, where R is the set of pixels inside the region of mass, and $|\cdot|$ is set cardinal.
2. Mass perimeter length. The perimeter length P is the total length of the mass edge. The mass perimeter length was computed by finding the boundary of the mass, then counting the number of pixels around the boundary.
3. Compactness. The compactness C is a measure of contour complexity versus enclosed area, defined as:

$$C = \frac{P^2}{4\pi A} \quad (1)$$

where P and A are the mass perimeter and area respectively. A mass with a rough contour will have a higher compactness than a mass with smooth boundary.

4. Normalized radial length. The normalized radial length is sum of the Euclidean distances from the mass center to each of the boundary coordinates, normalized by dividing by the maximum radial length.
5. Minimum and maximum axis. The minimum axis of a mass is the smallest distance connecting one point along the border to another point on the border going through the center of the mass. The maximum axis of the mass is the largest distance connecting one point along the border to another point on the border going through the center of the mass.
6. Average boundary roughness.
7. Mean and standard deviation of the normalized radial length. The mean μ and standard deviation σ of the normalized radial length are computed as

$$\mu_i = \frac{1}{n} \sum_{k=1}^n R_k \quad \sigma = \sqrt{\frac{1}{n} \sum_{k=1}^n (R_k - \mu_i)^2} \quad (2)$$

8. Eccentricity. The eccentricity characterizes the lengthiness of a ROI. To this purpose a symmetric matrix A is defined as follows

$$A_{11} = \sum_{i=1}^N (x_i - X_0)^2, \quad A_{22} = \sum_{i=1}^N (y_i - Y_0)^2 \quad (3)$$

$$A_{12} = A_{21} = \sum_{i=1}^N (x_i - X_0)(y_i - Y_0) \quad (4)$$

where N is the number of the ROI pixels; x_i and y_i are the coordinates of a generic pixel, X_0 and Y_0 are the coordinates of the geometric center of the ROI. If λ_1 and λ_2 are the eigenvalue of the A matrix, in the elliptical approximation of the ROI region, the semi-axis values will be

$$S_1 = \sqrt{\left| \frac{\lambda_1}{2} \right|} \quad S_2 = \sqrt{\left| \frac{\lambda_2}{2} \right|} \quad (5)$$

Then the eccentricity is given by

$$eccentricity = \frac{S_1}{S_2} \quad (6)$$

with $S_1 < S_2$. An eccentricity close to 1 denotes a ROI like a circle, while values close to zero mean more stretched ROIs.

9. Roughness. The roughness index was calculated for each boundary segment (equal length) as

$$R(j) = \sum_{k=j}^{L+j} |R_k - R_{k+1}| \quad (7)$$

for $j = 1, 2, \dots, \frac{n}{L}$ where $R(j)$ is the roughness index for the j -th fixed length interval.

10. Average mass boundary. The average mass boundary calculated as averaging the roughness index over the entire mass boundary

$$R_{ave} = \frac{L}{n} \sum_{j=1}^{\frac{n}{L}} R(j) \quad (8)$$

where n is the number of mass boundary points and L is the number of segments.

11. Region-based shape descriptor utilizes a set of moments. A small set of lower order moments is used to discriminate among different images. The most common moments are the geometrical moments, the central moments and the normalized central moments and the moment invariants.

$$\phi_1 = \mu_{20} + \mu_{02}$$

$$\phi_2 = [\mu_{20} - \mu_{02}]^2 + 4\mu_{11}^2$$

$$\phi_3 = [\mu_{30} - 3\mu_{02}]^2 + [3\mu_{21} - \mu_{03}]^2$$


$$\phi_4 = [\mu_{30} + \mu_{12}]^2 + [\mu_{21} + \mu_{03}]^2$$

$$\phi_5 = [\mu_{30} - 3\mu_{12}][\mu_{30} + \mu_{12}][(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] + [3\mu_{21} - \mu_{03}][\mu_{21} + \mu_{03}][3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \quad (9)$$

$$\phi_6 = [\mu_{20} - \mu_{02}][(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11}[\mu_{30} + \mu_{12}][\mu_{21} + \mu_{03}]$$

$$\phi_7 = [3\mu_{21} - \mu_{03}][\mu_{30} + \mu_{12}][(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] - [\mu_{03} - 3\mu_{12}][\mu_{21} + \mu_{03}][3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]$$

Table 1. Hu Moments for Image with Figure 2a

Hu Moments	Image with right side Figure 2a	Image with left side Figure 2a
		
ϕ_1	8.6E-004	1.5E-003
ϕ_2	2.8E-010	1.2E-008
ϕ_3	6.8E-015	7.0E-011
ϕ_4	2.3E-013	1.6E-010
ϕ_5	7.3E-027	1.5E-020
ϕ_6	-3.2E-018	5.2E-015
ϕ_7	5.7E-027	6.3E-021

The microcalcifications are grouped into clusters based on their proximity. A set of ten features was initially calculated for each:

1. Number of calcifications in a cluster
2. Total calcification area / cluster area
3. Average of calcification areas
4. Standard deviation of calcification areas
5. Average of calcification compactness
6. Standard deviation of calcification compactness
7. Average of calcification mean grey level
8. Standard deviation of calcification mean grey level
9. Average of calcification standard deviation of grey level
10. Standard deviation of calcification standard deviation of grey level.
11. Moments of the boundaries microcalcifications. Boundaries are characterized by an ordered sequence Euclidean distances between the centroid of the region microcalcification and all contour pixels. We have

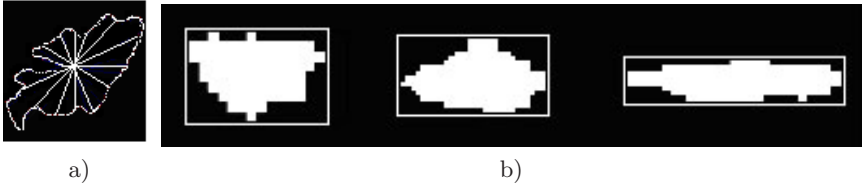


Fig. 3. Euclidean distances $b(i)$ a) and typical microcalcifications b)

$$M_1 = \frac{\left[\frac{1}{N} \sum_{i=1}^N \left(b(i) - \frac{1}{N} \sum_{i=1}^N b(i) \right)^2 \right]^{\frac{1}{2}}}{\frac{1}{N} \sum_{i=1}^N b(i)}, \quad M_2 = \frac{\left[\frac{1}{N} \sum_{i=1}^N \left(b(i) - \frac{1}{N} \sum_{i=1}^N b(i) \right)^4 \right]^{\frac{1}{4}}}{\frac{1}{N} \sum_{i=1}^N b(i)} \tag{10}$$

Moments M_1, M_2 are best feature to represent roughness of microcalcifications. The shape of microcalcifications are represented by maximum value of M_1, M_2 and mean of M_1, M_2 in each cluster.

The mass features together with features describing microcalcifications are the inputs of suitable modules of the CBIR systems.

3 Results and Conclusion

In this paper we presented microcalcifications and mass feature extraction in CBIR mammography system.

We used images from the MIAS digital mammography database which includes a description of the locations and types of the abnormalities [8].

The distance between two feature vectors is considered as a sum of distances relative to each of component pairs. We use the Bhattacharyya distance as the distance D_i between the component of the two feature vectors with $\forall i = 2, \dots, n$ where $n + 1$ is the size of the feature vectors.

Distance D between two feature vectors is chosen as

$$D = \sum_{i=0}^n D_i \tag{11}$$

Generally, given two images I_1, I_2 represented by feature vector set $Feat_1, Feat_2$ the similarity $d(Feat_1, Feat_2)$ is measured by sum of cross-correlation between the corresponding components

$$d(Feat_1, Feat_2) = \frac{1}{K} \sum_{k=1}^K \frac{Feat_1^k \cdot Feat_2^k}{\|Feat_1^k\| \|Feat_2^k\|} \tag{12}$$

where $Feat_1^k, Feat_2^k$ denote the feature vectors of the k th component of the image I_1, I_2 , respectively.

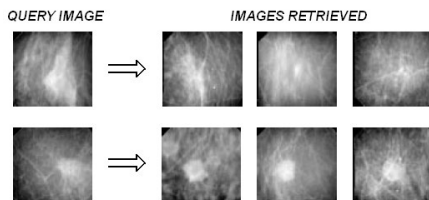


Fig. 4. Several matching results

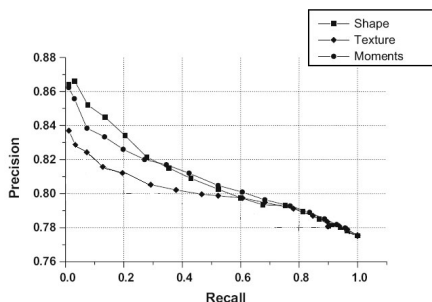


Fig. 5. Precision versus recall curve for various features of mass object

Figure 4 presents several matching results for mammograms with mass and Figure 5 presents curves of Precision vs. Recall obtained for shape, texture and moments features for mammograms with mass abnormalities.

Our future work include the design and development of an expert system for real time mammogram image analysis.

References

1. Highnam, R., Brady, M.: Mammographic Image Analysis. Kluwer Academic Publishers, Dordrecht (1999)
2. Linguraru, M., Brady, J.: A non-parametric approach to detecting microcalcifications. In: International Workshop on Digital Mammography. LNCS, Springer, Heidelberg (2002)
3. Shiffman, S., Rubin, G., Napel, A.: Medical Image Segmentation Using Analysis of Isolable-Contour Maps. IEEE Trans. on Medical Imaging 19, 1064–1074 (2000)
4. Zhang, M., Giger, M., Doi, K.: Mammographic texture analysis for the detection of spiculated lesions. In: Proc. 3rd International Workshop on Digital Mammography (1996)
5. Zwiggelaar, R., Parr, T., Schumm, J., Hutt, W., Taylor, J., Astley, S., Moggis, C.: Model-based detection of speculated lesions in mammograms. Medical Image Analysis 3, 39–62 (1999)

6. Chang, H.D., Shi, X.J., Min, R., Hu, L.M., Cai, X.P., Du, H.N.: Approaches for automated detection and classification of masses in mammograms. *Pattern Recognition* 39, 646–668 (2006)
7. Thangaval, K., Karnan, M., Sivakumar, R., Mohideen, K.A.: Automatic detection of microcalcification in mammograms - a review. *International Journal on Graphics, Vision and Image Processing* 5, 31–61 (2005)
8. <http://www.wiau.man.ac.uk/services/MIAS/>

Mathematical Morphology (MM) Features for Classification of Cancerous Masses in Mammograms

Konrad Bojar and Mariusz Nieniewski

Institute of Fundamental Technological Research, Polish Academy of Sciences,
Świętokrzyska 21, PL 00-049 Warsaw
kbojar@ippt.gov.pl, mnieniew@ippt.gov.pl

Summary. One of the important attributes of cancerous masses is their malignancy as it suggests a rapid growth of the cancer and possibility of metastasis. Malignancy, which denotes a special pathology of the tissue, is closely related to the existence of quasi-linear structures (spicules) emanating from the central mass. Hence, the tasks of malignancy and spicularity assessment are very often treated jointly. We propose a novel set of features enriching already existing pool of features for classification of masses. Our features are based on simple MM operations, pixel counting, and some basic algebra. To be more specific, given a contour of a cancerous mass we compute a sequence of dilations, and then count the number of pixels on the inner and the outer contour of each dilation. The contour pixel numbers are plotted against the size of the disk-shaped structuring element. The MM features are calculated from the plot via simple algebraic operations. The crucial point is that all the features are zero iff the input contour is circular. This distinctive property forms a basis for successful classification with the A_z values higher than for the features existing in the literature. The additional advantage of our approach is the simplicity of the proposed features. In contrast to many features found in the literature, no sophisticated algorithms are employed, so reimplementations of the features should be easy for anyone interested.

1 Introduction

The purpose of the screening mammography is the detection of breast cancer at an early stage of development. In contrast to the diagnostic mammography it involves evaluation of a large number of mammograms. The mammogram evaluation process may be divided into three parts: detection, or localization of masses, extraction of features from detected masses, and classification of masses into several possible classes. The main focus of the current paper is on the feature extraction for purposes of classification.

Let us assume that we are given a set of mammograms depicting cancerous masses identified by a skilled radiologist or by an automatic segmentation tool. Radiologists assign each of such masses several different attributes. Among these one finds malignancy and spicularity. Although it is a great simplification, throughout this paper we consider these attributes to be binary: a mass is regarded either malignant or benign, and spiculated or not spiculated. Up to now

many features for describing malignancy and spicularity have been proposed (see [1] for a comprehensive review). Just to mention a few, Fourier descriptors [8], acutance [7, 8], and SGLDM-based texture features [7, 10] are often used as measures of malignancy, while spicularity index [9], compactness [3, 9], fractional concavity [9], irregularity index [3] are measures of spicularity. Despite the fact that these features are reported by some authors to behave ideally on certain input data (for example, accuracy of 100% and 99.7% is reported in [5, 6]), their performance on data used in our analysis is rather poor (accuracy on the order of 55%-65%). It suggests that these features are quite fragile. More thorough analysis of the literature reveals that the above mentioned features can be computed by means of several different schemes, and each of these schemes leads to a different result. For example, polygonal approximation or derivatives of a curve can be computed in many ways, and there is no unique, canonical way to do that [2, 9]. Moreover, sometimes in various papers of the same author one finds various formulas for calculation of the same features [7, 8]. This illustrates the complexity of the classification task and demonstrates the urging need for further development in this area.

In this paper we concentrate on feature extraction putting aside the role of the classification scheme. We propose a novel set of features which can be used for the assessment of spicularity and malignancy. Although the new features taken by themselves do not assure high classification accuracy in the unsophisticated classification schemes, they are very simple and are built of common operations that are computed in unambiguous and precisely defined way. Hence, reimplementation of the features proposed is straightforward, and they can be easily incorporated into an existing pool of features (see [1] for an extensive review of features and classification schemes).

In Section 2 we introduce a set of novel features. In Section 3 we define some of the features already present in literature. In Section 4 we outline the details of conducted numerical experiments, and in Section 5 we draw conclusions from the conducted experiments.

2 Proposed MM Features for Mass Malignancy and Spicularity

The proposed set of malignancy and spicularity features is based on the morphological analysis of the shape of the mass only. Let M be a subset of the integer lattice \mathbb{Z}^2 . This set represents a binary image depicting the mask of the cancerous mass obtained from a skilled radiologist. Let D_n denote a digital disk of radius n with the origin in its center. Let \oplus denote the morphological dilation [4]. Let ∂ be the boundary operator given by the formula

$$\partial M = \{x \in M \mid \text{the set } (\mathbb{Z}^2 \setminus M) \cup \{x\} \text{ is 4-connected}\}, \quad (1)$$

where M is a mask of a mass. Let us define the *inner contour of index n* and the *outer contour of index n* of M by the respective formulas

$$C_{in}^n = \partial((\partial M \oplus D_n) \cap M) \setminus \partial M, \quad (2)$$

$$C_{out}^n = \partial((\partial M \oplus D_n) \cap (\mathbb{Z}^2 \setminus M \cup \partial M)) \setminus \partial M. \quad (3)$$

Using the inner and outer contours for a given $m = \pm n$ we define the *characteristic function* $L(m)$ to be

$$L(m) = \begin{cases} |C_{in}^{-m}| & m < 0 \\ |\partial M| & m = 0, \\ |C_{out}^m| & m > 0 \end{cases}, \quad (4)$$

where $|\cdot|$ is the cardinality operator. With N denoting the maximal radius of the disk, the characteristic function includes $2N+1$ points. The above definition was inspired by that of the pattern spectrum [4]. The $L(m)$ is similar to the pattern spectrum, but it does not have its properties. In particular, the pattern spectrum is defined by means of closing, whereas $L(m)$ is defined by means of dilation which is not idempotent. Moreover, closing with larger disks yields larger areas, whereas there is no such relation for the characteristic function $L(m)$, which does not have to be monotonic in its left or right branch (Fig. 2). However, the crucial point is that well circumscribed masses tend to have a linear characteristic function. And the more spiculated the mass is, the more characteristic function deviates from linearity (Fig. 2). Based on this remark we propose the following six auxiliary variables for the description of the characteristic function

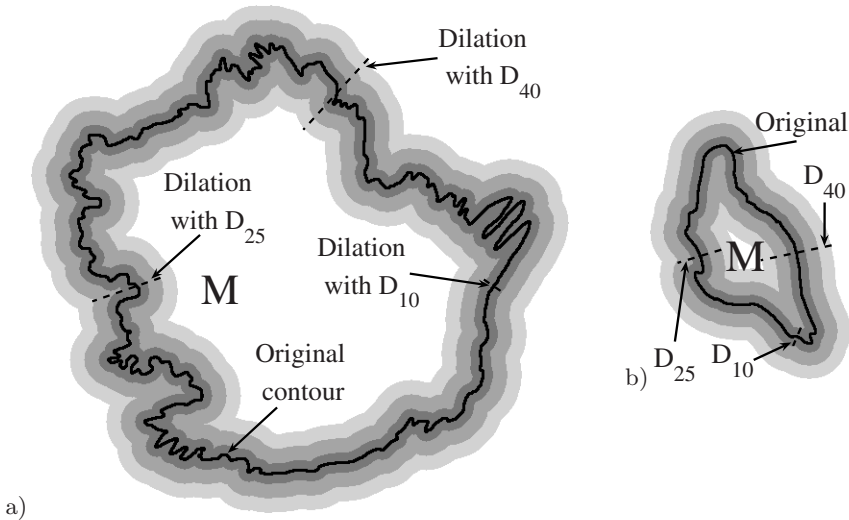


Fig. 1. Dilations of the contour of the masses: (a) MDB017LS and (b) MDB190RL from the MIAS database [11] by subsequent disks. The original contour, drawn by a skilled radiologist, was also dilated by cross-shaped structuring element D_1 strictly for visualization purposes. Relative dimensions of both masses are not preserved in this figure.

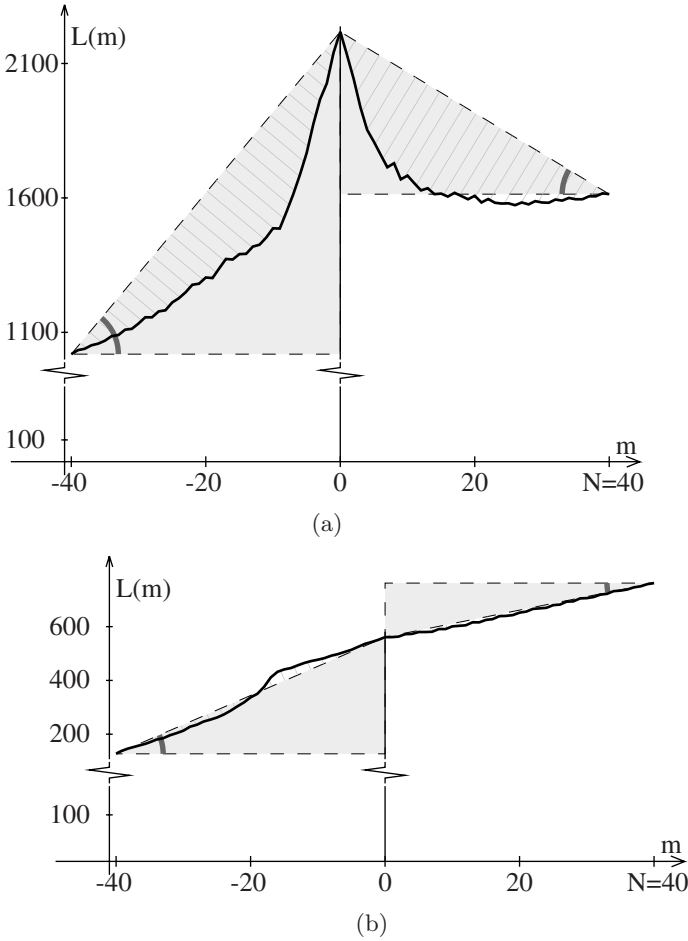


Fig. 2. The characteristic function (4) for the contours shown in Fig. 1. The left and the right plot correspond to the left and right contours, respectively. The angles at $m = -40$ and $m = 40$ are given by Eq. (5). The ratio of the left (or right) hatched area to the area of the left (or right) shaded triangle is given by Eq. (7).

$$\alpha_{in} = \text{atan} \left(\frac{L(0) - L(-N)}{N} \right), \quad \alpha_{out} = \text{atan} \left(\frac{L(0) - L(N)}{N} \right), \quad (5)$$

$$\Delta_{in} = 1 - \frac{L(-N)}{L(0)}, \quad \Delta_{out} = 1 - \frac{L(N)}{L(0)}, \quad (6)$$

$$R_{in} = \frac{\sum_{n=-N}^0 |L(n) - n \tan \alpha_{in} - L(0)|}{0.5N |L(0) - L(-N)|}, \quad R_{out} = \frac{\sum_{n=0}^N |L(n) + n \tan \alpha_{out} - L(0)|}{0.5N |L(N) - L(0)|}. \quad (7)$$

The variables (5) and (6) express the relation of the contour dilated by D_N to the initial contour (Fig. 2). The variables (7) measure the area contained between the graph of the characteristic function and the line connecting $L(0)$ and $L(N)$ or $L(-N)$, divided by the area of the respective shaded triangles in Fig. 2.

The variables (5)-(7) allow us to detect nonlinearities of the left and the right branch of the characteristic curve. For this purpose we construct the following MM features

$$\alpha = \alpha_{in} + \alpha_{out}, \quad (8)$$

$$\Delta = \Delta_{in} + \Delta_{out}, \quad (9)$$

$$R = R_{in} + R_{out}. \quad (10)$$

Note that all these features are close to zero for the circular contour. In the limiting case of the Euclidean plane these features would be exactly zero, and for elliptic and oval shapes they would be close to zero. Obviously, these features are rotation- and shift-invariant, but they are not scale invariant. Only α is normalized to the range $[-2\pi, 2\pi]$; and R can reach any non-negative value, while the value of Δ can be any real number.

3 Existing Features for Mass Malignancy and Spicularity

In this Section we define some of the malignancy and spicularity features mentioned in the introduction. We focused on the features referring to the contour together with its margins. We introduce compactness [3, 9], fractional concavity [9], spicularity index [9], and acutance [8]. The order of the features is chosen in accordance to their ascending complexity.

Before going to particular definitions we introduce some general notation. We denote by $L(0)$ the number of pixels of the initial contour ∂M (Eq. (4)), and i is used as an indexing variable along the curve. We denote the pixels of the contour by $p(i) = (x(i), y(i))$, where x and y are the coordinates of p at the i -th pixel. Every pixel on the contour p is m -adjacent to a single previous pixel and to a single subsequent pixel. Given a natural number K we define the derivative of p of size K at the i -th pixel by the formula

$$p'_K(i) = (x'_K(i), y'_K(i)) = \frac{1}{K} \sum_{k=1}^K \frac{p(i+k) - p(i-k)}{2k}. \quad (11)$$

However, our computations indicated that the final results are almost independent of K for $K \geq 5$. For this reason we drop the subscript and put $K = 5$ by default. Note that the contour under consideration is closed, so that p is periodic and $p(|\partial M| + k) = p(k)$. As a result, there are no boundary problems in Eq. (11). We also denote the sign of the curvature of p at any pixel by the symbol

$$\tilde{\kappa} = \text{sign}(x'y'' - x''y'). \quad (12)$$

Let us consider a sequence of all inflection points $p(j_1), \dots, p(j_J)$, that is the points in which $\tilde{\kappa}$ changes. Then, the parts of the contour between consecutive

inflection points are further divided as described in [9, Sec. 2.2]. The division points $p(i_1), \dots, p(i_I)$ determine the polygonal approximation of p . The above sequence is also periodic, and $p(i_{I+k}) = p(i_k)$. Let us denote the segment connecting the points $p(i_k)$ and $p(i_{k+1})$ by I_k . In the following $\|I_k\|$ denotes the Euclidean length of I_k .

The definitions of the features can now be formulated as follows.

Compactness CO. The compactness is a simple measure of circularity of the contour ∂M and is given by

$$CO = 1 - 4\pi|M|/\text{diameter}^2(M). \tag{13}$$

This feature is zero iff M is a disk, and grows towards unity with rising complexity of ∂M .

Fractional concavity FC. The fractional concavity is given by the Euclidean length of all concave segments in the polygonal approximation of p divided by the total Euclidean length of all segments [9, Sec. 3.1]

$$FC = \frac{\sum_{\substack{k \in \{1, \dots, I\} \\ \forall l \in \{i_k, \dots, i_{k+1}\} \tilde{\kappa}(l) \leq 0}} \|I_k\|}{\sum_{k \in \{1, \dots, I\}} \|I_k\|}. \tag{14}$$

This feature is zero for convex contours, and grows with contour oscillations. Unfortunately, this measure is fragile to small variations of almost linear contour parts.

Spicularity index SI. A spicule is defined as a sequence of segments $I_{i_k}, \dots, I_{i_{k+w}}$, with $w \geq 1$, between a pair of consecutive inflection points $p(j_k), p(j_{k+1})$. If $w = 1$, then the spicule is merged with its neighbor [9, Sec. 3.2], so we can assume that every spicule consists of at least two segments ($w \geq 2$). Let $\Theta_1, \dots, \Theta_{w-1}$ be the angles between consecutive segments, and let $\Theta_{th} = (\Theta_1 + \dots + \Theta_{w-1})/(w - 1)$ be the average angle. Given the spicule narrowness $\theta_k = \sum_{\{i \mid \Theta_i \leq \Theta_{th}\}} \Theta_i / |\{i \mid \Theta_i \leq \Theta_{th}\}|$, and the spicule length $S_k = \|I_{i_k}\| + \dots + \|I_{i_{k+w}}\|$, we calculate the spicularity index by the formula

$$SI = \frac{\sum_{k=1}^J (1 + \cos \theta_k) S_k}{\sum_{k=1}^J S_k} = 1 + \frac{\sum_{k=1}^J \cos \theta_k S_k}{\sum_{k=1}^J S_k}. \tag{15}$$

The SI tends to zero for less curvilinear contours, and grows with the appearance of spicules. The SI can reach the maximum of 2. Unfortunately, this feature is sensitive to the order in which spicules are merged. Moreover, it is fragile to small contour variations. For example, adding a small and narrow protrusion to the contour may change the SI drastically.

Acutance AC. The acutance is widely known in signal theory. In image processing it is defined as follows. At first the bundle of normals to the contour p is computed: for every i and for all $k \leq K = 5$ the normals $W_k(i)$ to the

lines passing through the points $p(i - k)$ and $p(i + k)$ are calculated. Then, the average normal $W(i)$ at $p(i)$ is a line passing through $p(i)$ and satisfying the relation $\angle(W(i), X \text{ axis}) = \frac{1}{K} \sum_{j=1}^K \angle(W_j(i), X \text{ axis})$. Subsequently, for each average normal $W(i)$ the Euclidean distance $v(i)$ from $p(i)$ to the nearest point different from $p(i)$ and in which $W(i)$ intersects p is computed. The average normal $W(i)$ is then clipped at $\lfloor v(i) \rfloor$ in the inner and at $\lceil v(i) \rceil$ in the outer direction yielding an interval with $p(i)$ in its center. In the next step, the gray-values $f(i, j)$ and $b(i, j)$ in the original mammogram are calculated at equidistant points in the inner and the outer directions, respectively, along clipped normals $W(i)$ using bilinear interpolation. Defining the average strength of the normal gradient at i -th pixel $d(i) = \sum_{j=1}^{\lfloor v(i) \rfloor} \frac{|f(i, j) - b(i, j)|}{2j}$ along $W(i)$, we calculate the acutance using the formula

$$AC = \frac{1}{\max_i d(i)} \sqrt{\frac{1}{|\partial M|} \sum_{i=1}^{|\partial M|} d^2(i)}. \quad (16)$$

The acutance is fragile to even smallest variations of the contour, both in high and low frequency, because the slightest variation of contour pixels may cause a large change in position of the endpoints of the clipped normals W , changing the grayvalues f and b completely.

4 Experiments

In order to test discrimination capabilities of our features we conducted several numerical experiments. The input data were taken from the MIAS database [11]. This database consists of 321 mammograms with 90 mammograms containing masses. Two out of these 90 mammograms contain two masses each, and the other two contain no visible mass according to the radiologist. In summary, we used all available 86 mammograms containing a single mass and never coming from the same patient. This means that our set consisted of 86 statistically independent cases with contours extracted by a skilled radiologist. It must be emphasized that we did not reject inconvenient cases to enhance the results. Our data set included 50 benign and 36 malignant cases. In other words, it included 67 non-spiculated and 19 spiculated cases. From our point of view the extracted masks are complicated as the radiologist took into account all suspected regions and spicules. In particular, quasi-linear structures, such as Cooper's ligaments and vessels influence the shape of the mass. It is quite possible that another radiologist would draw the mask of the mass differently, which means that there is no unique way of interpreting mammograms. Obviously, the chosen set of features should be robust with respect to modification of the contours drawn by various radiologists.

In Sections 2 and 3 we introduced three new features and quoted four existing in the literature. At the first step we tested discrimination capabilities of new and old features separately in order to compare their performance. For this purpose 22

Table 1. Classification results for various feature sets

Feature set	Classif. malignant/benign	Classif. spiculated/non-spiculated
	$A_z \pm \sigma_{A_z}$	$A_z \pm \sigma_{A_z}$
α	0.677 ± 0.057	0.736 ± 0.071
Δ	0.651 ± 0.059	0.716 ± 0.067
R	0.675 ± 0.057	0.542 ± 0.077
α, Δ	0.675 ± 0.057	0.712 ± 0.068
α, R	0.679 ± 0.057	0.721 ± 0.073
Δ, R	0.665 ± 0.058	0.710 ± 0.068
α, Δ, R	0.675 ± 0.057	0.715 ± 0.067
CO	0.540 ± 0.061	0.573 ± 0.067
FC	0.612 ± 0.060	0.521 ± 0.071
SI	0.676 ± 0.057	0.632 ± 0.069
AC	0.611 ± 0.061	0.539 ± 0.077
CO, FC	0.594 ± 0.060	0.517 ± 0.068
CO, SI	0.588 ± 0.062	0.577 ± 0.065
CO, AC	0.522 ± 0.061	0.507 ± 0.076
FC, SI	0.516 ± 0.062	0.548 ± 0.071
FC, AC	0.610 ± 0.060	0.531 ± 0.073
SI, AC	0.514 ± 0.062	0.542 ± 0.077
CO, FC, SI	0.513 ± 0.061	0.522 ± 0.069
CO, FC, AC	0.581 ± 0.061	0.509 ± 0.070
CO, SI, AC	0.518 ± 0.061	0.526 ± 0.074
FC, SI, AC	0.507 ± 0.062	0.520 ± 0.074
CO, FC, SI, AC	0.536 ± 0.061	0.517 ± 0.074
all seven features	0.575 ± 0.601	0.605 ± 0.062

feature sets, composed of α , Δ , R , and CO , FC , SI , AC , were chosen for experiments, including either new or old features (Table 1). In the second step we tested how the features would perform jointly, and we used all seven features at the same time (the last row in Table 1). In total $2 \times (22 + 1)$ numerical experiments were conducted: 23 for malignant vs. benign classification, and 23 for spiculated vs. non-spiculated classification. To classify masses we used the Fisher Linear Discriminant (FLD) [6, Sec. 3.3] which is one of the simplest non-trivial classifiers. For evaluation purposes the leave-one-out procedure was used in order to avoid necessity of having separate training and testing sets. For every feature set, the FLD generated a collection of real numbers, each number equal to the scalar product of the feature vector and a vector describing the hyperplane separating the classes. The obtained collection of numbers were input into the ROCKIT software (currently available at http://xray.bsd.uchicago.edu/kr1/KRL_ROC/software_index6.htm) for estimating the ROC curve of the classifier, the area A_z under this curve and its standard deviation σ_{A_z} . If the resulting classifier gave $A_z < 0.5$, we chose the classifier with the opposite sign of the normal to the hyperplane separating the classes.

Considering the malignant vs. benign classification column in Table 1 we see that the proposed features α , Δ , R give a fair classification result of $A_z \in [0.65, 0.68]$, while the similar result for CO , FC , SI , AC is on the average much

worse. In the latter group, the *SI* gives the highest result of 0.68. It can be seen that combining features in this group significantly lowers the A_z . This effect can be eliminated by employing a more advanced classifier. Finally, when using all seven features the obtained result was lower than in the case of proposed features only and higher than in the case of old features.

Considering the spiculated vs. non-spiculated classification column in Table 1 we remark that almost all combinations of features α , Δ , R give a classification result $A_z > 0.7$. Also here the *CO*, *FC*, *SI*, *AC* perform much worse giving $A_z = 0.63$ at most. In both feature groups adding more features to the set rarely enhances the result. In the case of all seven features, the result is slightly below the average of results of individual groups.

5 Conclusions

We proposed a set of new features (8)-(10) for discrimination between malignant vs. benign, and spiculated vs. non-spiculated masses in mammograms. The computation as well as coding of these features are much simpler than of features found in the literature. Moreover, the classification based on these features results in larger A_z values. Although this is not a proof of a higher discriminatory power as only one of the simplest classifiers was used, they already turned out useful in solving the classification problem under consideration. Further experiments using more mammographic databases will be performed, and additional features exploiting properties of the characteristic function (4) will be developed for even higher A_z .

Acknowledgement. The research was financed by the (Polish) Ministry of Education and Science as the research project No 3 T11C 050 29 in 2005-2008. We thank Prof. R. M. Rangayyan from the University of Calgary for his remarks and help in reimplementing of his works.

References

1. Cheng, H., Shi, X., Min, R., et al.: Approaches for automated detection and classification of masses in mammograms. Pat. Rec. 39, 646–668 (2006)
2. Fdez-Valdivia, J., García, J., de la Blanca, N., et al.: A new methodology to solve the problem of characterizing 2-D biomedical shapes. Comp. Meth. and Programs in Biomed 46, 187–205 (1995)
3. Lee, T., McLean, D., Atkins, M.: Irregularity index: A new border irregularity measure for cutaneous melanocytic lesions. Med. Image Anal. 7, 47–64 (2003)
4. Maragos, P.: Pattern spectrum and multiscale shape representation. IEEE Trans. on Pattern Anal. Mach. Intell. 11(7), 701–716 (1989)
5. Mu, T., Nandi, A., Rangayyan, R.: Analysis of breast tumors in mammograms using the pairwise Rayleigh quotient classifier. J. Electronic Imaging 16(4) (2007)
6. Mu, T., Nandi, A., Rangayyan, R.: Classification of breast masses via nonlinear transformation of features based on a kernel matrix. Med. Bio. Eng. Comp. 45, 769–780 (2007)

7. Mudigonda, N., Rangayyan, R., Desautels, J.: Gradient and texture analysis for the classification of mammographic masses. *IEEE Trans. on Med. Imaging* 19(10), 1032–1043 (2000)
8. Rangayyan, R., El-Faramawy, N., et al.: Measures of acutance and shape for classification of breast tumors. *IEEE Trans. on Med. Imaging* 19(6), 799–810 (1997)
9. Rangayyan, R., Mudigonda, N., Desautels, J.: Boundary modelling and shape analysis methods for classification of mammographic masses. *Med. Bio. Eng. Comp.* 38, 487–496 (2000)
10. Sahiner, B., et al.: Classification of mass and normal breast tissue: On convolution of neural network classifier with spatial domain and texture images. *IEEE Trans. on Med. Imaging* 15(5), 598–610 (1996)
11. Suckling, J., Parker, J., et al.: The Mammographic Images Analysis Society digital mammogram database. In: Gale, A.G., Astley, S.M., et al. (eds.) *Digital Mammography*. Excerpta medica int'l congress series, vol. 1069, pp. 375–378 (1994)

Stroke Display Extensions: Three Forms of Visualization

Artur Przelaskowski¹, Katarzyna Sklinda², and Grzegorz Ostrek¹

¹ Institute of Radioelectronics, Warsaw University of Technology Nowowiejska 15/19, Warszawa, Poland

arturp@ire.pw.edu.pl

² Department of Radiology CMKP, CSK MSWiA, Woloska 137, Warszawa, Poland

katarzyna.sklinda@gmail.com

Summary. The computer-assisted support of acute ischemic stroke detection was the subject of our research reported in this paper. The conditioning of early stroke diagnosis based on CT examinations was analyzed. The multiscale extraction of the subtlest signs of hypodensity which were often undetected in standard CT scan review was presented. Proposed method was as follows: evidence-based description of ischemic changes, the analysis of hypodensity signs across scales, noise reduction and hypodensity extraction, and following display of ischemic changes localized in source brain image space. Important issues were: –extension of the brain tissues for marginal and missing space after deskulling and segmenting of unusual areas; –multiscale transform selection; –denoising in scale-space domain; –visualization conditions fixing. Three forms of extracted stroke signs visualization were proposed. Increased visibility of cerebral ischemia for difficult-in-diagnosis cases was experimentally noticed.

1 Ischemic Stroke Conditioning

Stroke is a clinical manifestation of diverse pathologies, defined as a syndrome characterized by rapidly progressing clinical signs and/or symptoms of focal loss of cerebral function lasting more than 24 hours or leading to death, with cause of vascular origin. An average district general hospital admits approximately 300 - 400 stroke patients a year, one third of which dies within first three months and one third acquires significant long-term disability. The common awareness of stroke including how it should be managed is still poor and the medical care, including early radiological detection, has failed to keep pace with developments in its treatment [1].

Brain imaging is required to guide the selection of acute interventions to treat patients with a stroke, which is very important for stroke emergency centers. The clinical diagnosis of acute stroke is sometimes difficult and thus the role of neuroimaging is gaining significance. It should allow identification of patients with hyperacute infarctions and selection of treatment, assessment of penumbra (tissues at risk of infarction), determination of etiology and follow-up of therapy and its possible complications. For most cases, CT remains the most important

brain imaging test. However, new studies suggest that MRI may also be used to detect the acute intracerebral hemorrhage and that it could be an alternative to CT. Additional studies are under way [2].

Multiscale imaging of hyperacute stroke pathology enables early assessment of reversible ischemic injury [3]. On a CT scan it would be represented by a focal hypodense area, in cortical, subcortical, or deep gray or white matter. A hypodense area is defined as any area in the brain with density lower than normal surrounding brain tissues. Infarcted territory is usually either wedge-shaped or round, occupying a defined vascular territory. It would be described according to the location and size.

On the initial CT-scan, performed during the hyperacute phase of stroke, (0-6h) the mentioned hypodensity does not have to be seen. Early indirect findings, like obscuration of gray/white matter differentiation and effacement of sulci, or "insular ribbon sign", may be noticed instead. Afterwards, it becomes possible to detect a slight hypodense area of infarction either in the cortices or the basal ganglia. Initially, the low density region is poorly defined, becoming more sharply delineated in the following hours [3, 4, 5, 6, 7]. The recent advent of thrombolytic therapy for hyperacute stroke treatment makes the earliest detection of areas of hypoattenuating ischemic parenchyma exceedingly important [4, 5, 6].

1.1 Pathophysiology

Stroke can be divided into ischemic - resulting from infarction of the brain, and hemorrhagic - resulting from intracerebral/subarachnoid hemorrhage. First mention form is much more common as approximately 85 percent of strokes result from infarction. A constant supply of oxygenated blood is necessary to maintain proper function of central nervous system (CNS). Since, unlike other organs, the brain is unable to store energy, it also requires a constant supply of glucose. The adult human brain with a cerebral blood flow (CBF) of 800 ml/min (15-20% of the total cardiac output) uses about 20% of all body oxygen. It is a high flow, low pressure system with preserved flow during diastolic phase [9]. The lowest acceptable CBF level for the brain is 15-25 ml/100 g per minute. CBF below 10-15 ml/100 g per minute leads to death of the tissues. The brain tissues may survive in a state of dysfunction, if CBF value higher than 12 ml/100 g per minute. It is an important clinical information whether the ischemic brain has a chance of survival as it affects further treatment.

Stroke is may caused by many different mechanisms, but all come down to some disruption of CBF and subsequent tissue damage. In spite of well-developed collateral circulation, there are cells and regions of the brain that are more vulnerable to changed perfusion than the other. Ischemia affects firstly the neurons, then the astrocytes, oligodendroglia and microglia. Brain cortex water content increases immediately after arterial occlusion. Earliest changes visible within neurons concern their mitochondria which become swollen and disorganized. A failure of the ATP-driven Na/K pump leads to increased intracellular water uptake. Increasing concentration of glutamate affects calcium channels resulting in further increase of intracellular water levels. All of these mechanisms result

in cytotoxic edema. During the initial 3 h of ischemia, the intracellular increase in water and sodium contents is almost exclusively confined to the gray matter. Between 4 and 6 hours after onset of ischemia the neurons begin to shrink, the synaptic gaps enlarge with expansion of astrocytic end feet, which lasts up to 24 h after ischemia. Infarcted region is characterized by loss of the borders between white and gray matter and focal swelling with effacement of the gyri. This swelling, due to intracellular cytotoxic edema, provokes decrease of tissues density in CT, which reaches maximal values between 24 and 48 h after infarction. The reparative and resorptive mechanisms start about 24-48 h after, beginning at the periphery of the infarcted zone and proceeding towards its core.

The final stage of infarction is accomplished by 2-4 weeks after onset and may be recognized by gliosis, shrunken gyri, encephalomalacic cysts, enlarged sulci and adjacent dilatation of CSF-containing spaces, but resorption of the necrotic tissue may last for months.

1.2 CT Findings in Acute Stroke

Due to its availability, CT remains the "gold standard" for detection of cerebral hemorrhage and maintains a principal position in the evaluation of patients with acute stroke. Generally, there is need to examine a stroke patient with CT is as soon as possible, although many infarcts do not emerge on CT until hours after the onset of stroke. A typical sign of acute infarction on CT is hypodense zone within a defined arterial supply territory which is thought to correspond irreversibly damaged brain tissues. The majority of larger infarcts reveal within 6 h. If brain water content increases by 1%, CT attenuation decreases by 2-3 HU (Hounsfield Unit) [10]. Edema, which occurs initially in the ischemic cortex reduces its contrast with respect to the adjacent white matter and provokes a loss of anatomic margins. In the next stage, lasting for several days and up to 2 weeks so called "fogging" phenomenon occurs - infarcted zone becomes isodense and the extent of an infarction may be underestimated on CT. Then, by 2-3 months, the infarctions are easily visible as areas of water density.

Early CT findings in acute middle cerebral artery (MCA) infarction are obscuration of lentiform nucleus which is due to cellular edema in the basal ganglia and a hyperattenuating MCA representing acute thrombus within the M1 segment as well as loss of the gray/white interface at the lateral margin of the insula, and called it the "insular ribbon" sign described by Truwit et al. [11]. Effacement of the sulci may also occur in the hyperacute/acute stage as an effect of rising focal edema.

Lately, perfusion CT has proved to be beneficial in the evaluation of acute stroke patients as it allows for fast qualitative and quantitative assessment of cerebral perfusion by generating maps of CBF, cerebral blood volume (CBV), mean transit time (MTT) and time-to-peak (TTP).

1.3 Hypodensity Models

Focal hypodense changes were found to be the most frequent and reliable signs of early cerebral ischemia. A decline in cerebral blood flow causes the brain tissue to take up water immediately. Thus, in the early stage of cerebral ischemia, the tissue changes consist mainly in alteration of water and electrolyte content. Parallel intracellular increase of sodium and a decrease of potassium concentration occur. A 2-4% increase in brain tissue water within 4 h of MCA occlusion was noticed in several experiments [3]. Increase of water content causes the lowering of brain attenuation coefficients in hyperacute ischemia, which leads to a discrepant decrease of about 1.3-2.6 HU for 1% change in water content [4]. The discrepancy of water uptake and density changes might suggest an incompleteness of ischemic physiology model and unclear impact of other factors, e.g. decreased lipids, increased protein and electrolyte changes.

According to standard templates, the site of the infarct is defined as subcortical, when internal border zone or deep arterial branch areas are involved, with additional differentiation between nonlacunar and lacunar infarcts. The latter is defined as a subcortical sharply delineated focal lesion with a diameter equal or less than 15 mm. Cortical infarct occurs, when superficial arterial branch territories are involved; cortico-subcortical, when concomitant involvement of arterial deep and superficial territories is present. The size of the infarct is quantified as follows - small, when the lesion involves less than one half a lobe (or in some studies when CT scans are permanently negative); medium, when the lesion involves one half to one lobe; and large, when the lesion involves more than one lobe. Mass effect is defined as slight, when only a compression of ventricles without dislocation is present; moderate, when a partial ventricular shift across the midline is observed; and severe, when a total ventricular shift across the midline is described. The age of the lesion, and hence its relevance to the present clinical symptoms, is usually judged from the degree of the mass effect, the clarity of its margins, the degree of hypodensity and presence of hemorrhagic transformation

Furthermore, the attenuation coefficients of brain parenchyma vary, mainly due to the differing thickness of the cranial vault. Dense bone lowers the energy of the beam and thus, increases attenuation. M. Bendszus et al. [12] found inter-individual differences, i.e. bone artifacts, of up to 14 HU in brain parenchyma at comparable scan levels. The sufficient accuracy, stability and linearity of CT number (HU) and degradation of contrast resolution caused by noise, are the next problem. The CT number for water should ideally be zero, but the actual value changes because of variations in the stability of the detector system or x-ray source. Normally, these variations (i.e. standard deviation of the water value) are very small and most scanners should be able to stay within 2 HU of zero for water. The mean CT number measured over the central test ROI (region of interests) should be in the range of 4 HU, which is close to the possible changes within acute ischemic region.

Nevertheless, it is evident that the early changes with ischemia occur, but may vary within the limited range of HU scale depending on cerebral infarct case, discrepant patient characteristics, and acquisition conditioning. The hypodense changes are slight, and ischemic area is not well outlined or contrasted (with slow edges characterized by low-frequency spectrum). Because of the human eye limitations, these first ischemic signs can often be out of that range. Typical preview window of width 80 HU gives maximum noticeable change of 1-2 grey shade within the first 4 h of ischemia. Diffusely interspersed changes in grey shade can hardly be distinguished in noisy areas because of a low brightness contrast, bone artifacts, non-optimum scanning.

The purpose of computer assistance was to improve the diagnosis of hyperacute ischemic brain parenchyma by enhanced hypodensity visualization on emergency CT scans. Suggested processing method was based on multiscale image analysis, region of interest segmentation and local signal extraction and visualization. Proposed method was based on the algorithm presented in [17] extended to enhanced multiscale analysis resulting in three visualization forms.

2 Materials and Methods

The proposed method was based on a concept of Stroke Display implied as intelligent data visualization method that communicates extracted and enhanced hypodensity to the observers, especially for "radiologically silent" cases (really difficult to diagnose). It complements conventional CT scan view with highly specific to infarct cases display [17]. As computer-assisted interpretation tool Stroke Display was designed to uncover, model and enhance signatures of ischemia. Multiscale transformations was used to analyze image content basing on spatially distributed soft tissue properties over different scales and subbands. Hypodensity may be effectively strengthened, extracted and identified through hierarchical local data processing. Post-processing in wavelet domain was less susceptible to local perturbations, and beneficial noise suppression and selective contrast enhancement was possible [13, 14, 15]. Especially, wavelet-based algorithms with adaptive histogram equalization were investigated as a method of automatic simultaneous display of the full dynamic contrast range of CT images (chest exams with lymph nodes, pericardial disease, air cysts) [16].

Generally, the algorithm steps are as follows: –segmentation of potentially hypodense areas (e.g. sulci or the aged lesions); –extension of the brain tissues for marginal and missing space after deskulling and segmenting of unusual areas; –multiscale transformation; –denoising in scale-space domain; –visualization of processed CT scans.

Initial two-stage segmentation of the regions susceptible to ischemic density changes was used to eliminate false diagnostic indications. Soft tissue areas with low density were fulfilled with higher density tissue properties and processed for final visualization. False positives have to be avoided, since treating ineligible patients with intravenous thrombolysis is associated with an unacceptable

risk of hemorrhage and death. Instead of image wavelet decomposition used in [17], curvelets (much better describing curves) [18] and mixed curvelet-wavelets processing was used for more effective noise suppression and more convincing regular hypodense areas visualization.

2.1 Algorithm of Stroke Display

The successive steps are as follows:

1. Segmentation of diagnostic ROIs
 - the brain extraction to remove non-brain tissue from a CT volume (to deskull the brain in the image) through region growing, arranged in 3D space of successive slices; initial set of seeds were grown-up to regular, spatial regions of the brain tissue identified in successive slices; fast implementation was optimized;
 - selection of the only tissue regions which are susceptible to ischemia called diagnostic ROIs; clear brain sulci, old ischemic scars and other structures useless in acute stroke detection were discarded; mixed growing-thresholding method was applied: –weighted average filtering for noise reduction, –detection of low density seeds based on adaptive thresholding, –growing low density regions with adaptive membership function, –correction of indicated areas to make them regular, smooth and big enough sized;
 - complement of diagnostic ROIs with mean values of neighbor areas providing the continuity of density function and absence of lower density regions;
2. Hypodensity extraction
 - multiscale decompositions; three forms were used for different processing effects: –curvelets (according to FDCT implementation [19]) for perception improvement, –curvelets followed by non-perfect orthogonal wavelet base tspline2 (kernel defined by low pass analysis filter $\tilde{h} = [1/4, 2/4, 1/4]$) for soft detector indications, – tspline2 followed by curvelets for clear detector indications;
 - adaptive soft thresholding of curvelet coefficient modulus for subtle denoising and increasing the local mean data variability;
3. Visualization of diagnostic image content
 - 8 bit display arrangement with contrast enhancement by adaptive histogram equalization of processed and synthesized image data; the brain areas processed in multiscale domain (according to three decomposition forms) are reconstructed, adaptively converted to suitable presentation scales and fitted to the source image with skull, scalp and surrounding tissue in the best view window;
 - 3 forms of final visualization based on and adjusted to 3 decomposition and processing methods; considering progressive stroke signs extraction we have: –MU-PP for improved perception of density distinctions, –MU-DD for softly indicated hypodensity areas, –MU-DE for clearly indicated hypodensity.

3 Experimental Study

Preliminary subjective tests were performed to verify the efficiency of extended Stroke Display implementation. The valid presence and indicated position of asymmetric hypodense signs compared to the reference diagnosis (the location and size of the infarct) based on follow-up CT scans (from 1 to 10 days after the ictus) were used as performance criteria. According to simple test rules and clear visualization of extracted hypodense areas, display usefulness could be simply

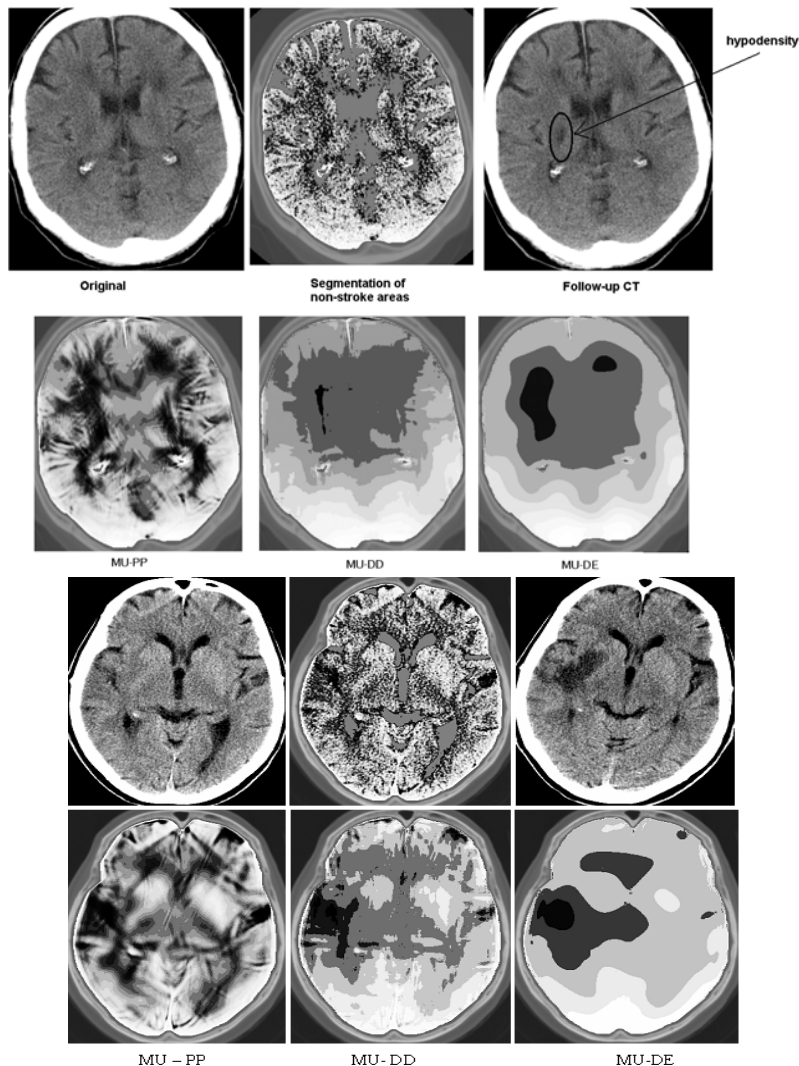


Fig. 1. Different forms of visualization (MU-PP, MU-DD, MU-DE) in Stroke Display; presented cases were used in the experiments

Table 1. The experimental results of extended Stroke Display verification compared to the reference results reported in [17]; mean scores of 9 observers for 6 cases difficult in diagnosis were presented

Visualization	Precision	Sensitivity	Specificity	PVP
Original - [17]	0.67	0.58	0.75	0.5
MU - [17]	0.42	0.5	0.33	0.38
MU - PP	0.56	0.52	0.52	0.49
MU - DD	0.54	0.59	0.48	0.56
MU - DE	0.80	0.89	0.70	0.74

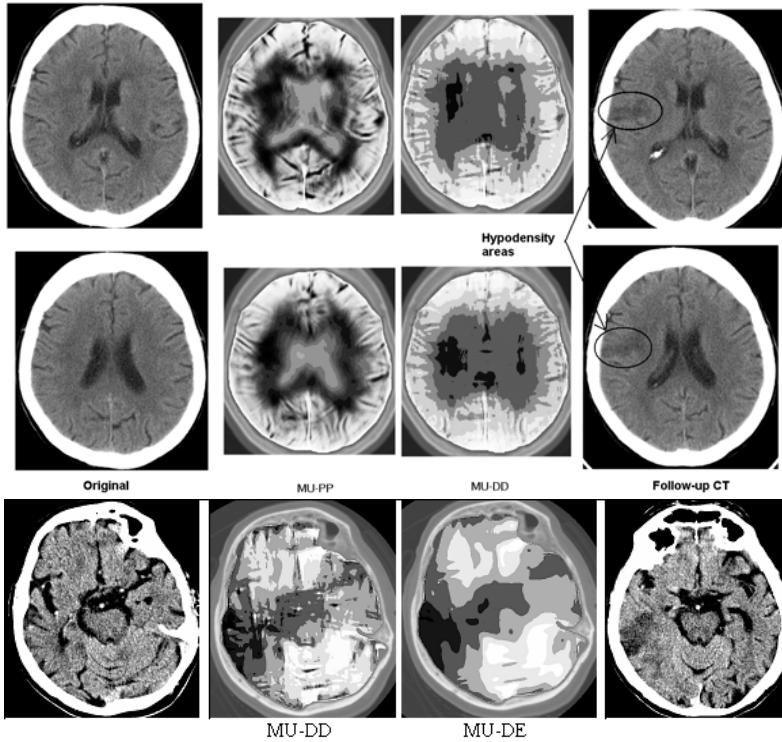


Fig. 2. Other test cases of Stroke Display assistance for ischemia diagnosis procedure

verified even by non-specialists, biomedical engineers or medicine students with the aim of preliminary assessment of Stroke Display extended to 3 additional forms of visualization. 9 observers participated in the experiments: radiologist, 4 medical students and 4 biomedical engineers or students to confirm clearness of display indications.

A test set consisted of 6 CT exams of brain selected as extremely difficult in diagnosis basing on false decision of specialists and unusual role of Stroke Display in the experiments reported in [17]. For 3 clinically confirmed cases of acute stroke appearance the time between the onset of symptoms and the early CT examination was ranged from 1 to 5 hours. 3 other were clinically confirmed normal cases. The examples of different display visualization forms were presented in Fig. 1 and 2. The clearness of hypodensity extraction and any false suggestions dominantly affected observer responses in ischemic stroke detection procedure. The results were presented in Table 1.

According to observers opinions, extended Stroke Display improved the diagnosis for difficult cases because of distinct visibility of hypodense signs or generally ischemic susceptibility for most cases of test examinations. Clear indications of MU-DE increased significantly stroke detection efficiency according to several measures in comparison to typical interpretation of originals, poorly effective MU version from [17] and less expressive forms of extended Stroke Display. Resulting higher efficiency of diagnosis for non-specialists assisted with additional display in comparison to standard diagnosis of specialists is surprising and confirms usefulness of extended Stroke Display.

4 Conclusions

Presented results indicate that hypodensity-oriented perception improvement and detection of ischemic areas may facilitate the interpretation of CT scans in hyperacute infarction. Effective multiscale extraction of pathology signs with different forms of content visualization extends available computer assistance tools in a way that was useful and significantly increased efficiency of diagnosis, especially for difficult cases of ischemia.

Therefore, the reliable computer assistance with expressive display of distinct ischemic signs can considerably improve the diagnosis of hyperacute ischemic stroke with increased objectivity, observer conviction, decision sensitivity and high enough specificity. Further optimization of computer-based understanding of stroke phenomenon is possible and desired. Clinical tests are necessary to consider the display as possible to be accepted for medical practice.

References

1. Rudd, A., Olfe, C.: Aethiology and pathology of stroke. *Hospital Pharmacist* 2, 32–36 (2002)
2. Adams, H., Adams, R., Del Zoppo, G., Goldstein, L.B.: Guidelines for the early management of patients with ischemic stroke, 2005 update. A scientific statement from the Stroke Council of the American Heart Association/American Stroke Association. *Stroke* 36, 916–921 (2005)
3. von Kummer, R.: The impact of CT on acute stroke treatment. In: Lyden, P. (ed.) *Thrombolytic Therapy for Stroke*, Humana Press, Totowa, New Jersey (2005)
4. Tomura, N., Uemura, K., et al.: Early CT finding in cerebral infarction. *Radiology* 168, 463–467 (1988)

5. Bozzao, L., Bastianello, S., et al.: Correlation of angiographic and sequential CT findings in patients with evolving cerebral infarction. *AJNR Am. J. Neuroradiol.* 10, 1215–1222 (1989)
6. von Kummer, R., Allen, K.L., et al.: Acute stroke: usefulness of early CT findings before thrombolytic therapy. *Radiology* 205, 327–333 (1997)
7. Wardlaw, J.M., Mielke, O.: Early signs of brain infarction at CT: observer reliability and outcome after thrombolytic treatmentsystematic review. *Radiology* 235, 444–453 (2005)
8. <http://www.strokeassociation.org/presenter.jhtml?identifier=1020>
9. Markus, H.S.: Cerebral perfusion and stroke. *J. Neurol. Neurosurg. Psychiatry* 75, 353–361 (2004)
10. Unger, E., Littlefield, J., Gado, M.: Water content and water structure in CT and MR signal changes: possible influence in detection of early stroke. *Am. J. Neuroradiol.* 9, 687–691 (1988)
11. Truwit, C.L., Barkovich, A.J., Gean-Marton, A., Hibri, N., Norman, D.: Loss of insular ribbon: another early CT sign of acute middle cerebral artery infarction. *Radiology* 176, 801–806 (1990)
12. Bendszus, M., Urbach, H., et al.: Improved CT diagnosis of acute middle cerebral artery territory infarcts with density-difference analysis. *Neuroradiology* 39(2), 127–131 (1997)
13. Starck, J.L., Murtagh, F., Candes, E.J., Donoho, D.L.: Gray and color image contrast enhancement by the curvelet transform. *IEEE Trans. Image Proc.* 12(6), 706–717 (2003)
14. Bonnier, N., Simoncelli, E.P.: Locally adaptive multiscale contrast optimization. *Proc. IEEE ICIP* 2, 1001–1004 (2005)
15. Hammond, D.K., Simoncelli, E.P.: Nonlinear image representation via local multiscale orientation. Courant Institute Technical Report TR2005–875 (2005)
16. Fayad, L.M., Jin, Y., et al.: Chest CT window settings with multiscale adaptive histogram equalization: pilot study. *Radiology* 223, 845–852 (2002)
17. Przelaskowski, A., Sklinda, K., Bargiel, P., Walecki, J., Biesiadko-Matuszewska, M., Kazubek, M.: Improved early stroke detection: wavelet-based perception enhancement of computerized tomography exams. *Comp. Biol. Med.* 37, 524–533 (2007)
18. Candes, E.J., Donoho, D.: Curvelets - a surprisingly effective nonadaptive representation for objects with edges. In: Schumaker, L.L., et al. (eds.) *Curves and Surfaces*, Vanderbilt University Press, Nashville (1999)
19. Candes, E.J., Demanet, L., Donoho, D., Ying, L.: Fast discrete curvelet transforms. *Multiscale Model Simul.* 5, 861–899 (2005)

Automated Fuzzy-Connectedness-Based Segmentation in Extraction of Multiple Sclerosis Lesions

Jacek Kawa and Ewa Pietka

Silesian University of Technology, Department of Biomedical Engineering, Gliwice,
Poland
jkawa@polsl.pl

Summary. In the current study, a fuzzy-connectedness-based approach to fine segmentation of demyelination lesions in Multiple Sclerosis is introduced as an enhancement to the existing ‘fast’ segmentation method. First a fuzzy connectedness relation is introduced, next a short overview of the ‘fast’ segmentation method is presented. Finally, a novel, automated segmentation approach is described. The combined method is applied to segmentation of clinical Magnetic Resonance FLAIR Images.

1 Introduction

Multiple sclerosis (MS) is an inflammatory demyelinating disease of the central nervous system. It is characterised by multiple plaques of demyelination in the white matter of the brain and spinal cord.

An automated segmentation of Multiple Sclerosis (MS) demyelination plaques in Magnetic Resonance (MR) images is a subject of many studies. Algorithms for a segmentation of the normal and abnormal white matter [1], as well as segmentation of lesions in MS [2, 3, 4] employ supervised and automated, both fuzzy and non-fuzzy approaches.

In [5], a kernel-space fuzzy c-means clustering method is used for the segmentation of plaques of the Fluid Light Attenuation Inversion Recovery (FLAIR) MR images. The approach employs features extracted from the entire MR volume. This shortens the processing time, yet the method tends to undersegment lesions in the presence of local inhomogeneities. In the current study, an enhancement to the ‘fast’ method is presented, based on the fuzzy connectedness concept [6].

2 Fuzzy Connectedness

A fuzzy connectedness (FC) is a fuzzy 2-ary relation in the image spel set. The axiomatic definition [6] bases on the 2-ary affinity, adjacency, and κ -net fuzzy relations defined within a fuzzy digital scene on \mathbf{Z}^n space. In this paper a simplified, graph-based view for \mathbf{Z}^2 is used. Let $\mathbf{z}_i = (z_{i1}, z_{i2})$ denote an image

pixel with the signal level $f(\mathbf{z}_i)$ and let all the image pixels \mathbf{z}_i constitute nodes of a graph.

In the graph, each pixel (each node) is interconnected with (and only with) all the pixels that are within its spatial neighbourhood (e.g. only 9 or 25-connected image pixels have links between them within the graph). Every direct \mathbf{z}_i - \mathbf{z}_j link in the graph has a strength assigned to it, that is equal to the value of a reflexive and symmetric fuzzy affinity relation for the two connected nodes $\mu_{\kappa_v}(\mathbf{z}_i, \mathbf{z}_j)$.

Within the graph, for each two pixels $\mathbf{z}_k, \mathbf{z}_l$, a path can be found that consist of zero or more links. A set of all existing paths from \mathbf{z}_k to \mathbf{z}_l is denoted as $P_{z_k z_l}$.

To each path p from $P_{z_k z_l}$, the strength $\mu_{\mathcal{N}}(p)$ is assigned as the strength of its weakest link, i.e. the lowest value of affinity for two constituting nodes ($[0, 1]$). The fuzzy connectedness can be then defined as a fuzzy relation:

$$\forall \mathbf{z}_k, \mathbf{z}_l: \mu_K(\mathbf{z}_k, \mathbf{z}_l) = \max_{p \in P_{z_k z_l}} [\mu_{\mathcal{N}}(p)]. \quad (1)$$

The notion of local pixel similarity is therefore analytically defined by means of a membership function of the fuzzy affinity relation μ_{κ_v} . In segmentation tasks, the fuzzy affinity usually [6, 7] depends on the signal level of compared pixels, the local signal gradient, and properties of the extracted object – defined by additional parameters:

$$\forall \mathbf{z}_k, \mathbf{z}_l: \mu_{\kappa}(\mathbf{z}_k, \mathbf{z}_l) = \begin{cases} w \cdot G_1(\mathbf{z}_k, \mathbf{z}_l) + (1 - w) \cdot G_2(\mathbf{z}_k, \mathbf{z}_l), & \\ \quad \text{iff } \mathbf{z}_k \neq \mathbf{z}_l \text{ and } \mathbf{z}_k, \mathbf{z}_l \text{ are neighbours} & \\ 1 & \text{iff } \mathbf{c} = \mathbf{d}, \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where $w \in [0, 1]$, and G_1, G_2 are selected from:

$$g_1(\mathbf{z}_k, \mathbf{z}_l) = \exp \left(-\frac{1}{2s_1^2} \left(\frac{f(\mathbf{z}_k) + f(\mathbf{z}_l)}{2} - m_1 \right)^2 \right), \quad (3)$$

$$g_2(\mathbf{z}_k, \mathbf{z}_l) = \exp \left(-\frac{1}{2s_2^2} (|f(\mathbf{z}_k) - f(\mathbf{z}_l)| - m_2)^2 \right), \quad (4)$$

$$g_3(\mathbf{z}_k, \mathbf{z}_l) = 1 - g_1(\mathbf{z}_k, \mathbf{z}_l), \quad (5)$$

$$g_4(\mathbf{z}_k, \mathbf{z}_l) = 1 - g_2(\mathbf{z}_k, \mathbf{z}_l), \quad (6)$$

for $f(\mathbf{z}_k)$ and $f(\mathbf{z}_l)$ denoting signal level of \mathbf{z}_k and \mathbf{z}_l , respectively.

2.1 Object Segmentation

Given the similitude [6] relation of fuzzy connectedness, for any threshold t , an image may be divided into a number of distinct components, that are connected with $\mu_K \geq t$. By selecting a single seed point (pixel), the component including it is extracted. Therefore, if at least one point is selected, an object extraction can be performed¹. The direct segmentation method is rarely used, though, in compare to the relative fuzzy connectedness approach.

¹ In this case, if a greater number of seed points is specified, the segmentation object is composed of multiple image components.

For segmentation based on relative fuzzy connectedness, at least two seed points have to be selected: (1) one belonging to the extracted object, and (2) another belonging to the background (i.e. *not* a segmented object). The object is then extracted as the set of points for which the fuzzy connectedness with respect to the object seed point (points) is higher than the fuzzy connectedness computed against the background seed point (points).

3 Initial Conditions - Fast Segmentation Overview

The fine – fuzzy-connectedness-based MS lesion segmentation is preceded by a fast segmentation procedure presented in [5]. During the processing, a kernelized clustering method is used [8, 5], based on standard Fuzzy *c*-Means algorithm [9].

The clustering results of MR FLAIR series (Fig. 1a) in Gaussian kernel space are analysed in the histogram context and used during the initial classification of the brain tissue. The obtained binary masks (Fig. 1b-d) match the region of: (1) Cerebrospinal Fluid – CSF and background, (2) brain tissue, and (3) demyelination plaques and fat. Next, a morphology-based processing is applied. The intracranial mask is produced for each MR slice (Fig. 1e). Then, the eyes, and brain ventricles are detected and a CSF mask is created (Fig. 1f). After that step, another mask is produced (Fig. 1g) that matches the grey and white matter (including demyelinated tissue), and combined with the initial classification mask, permits the MS plaques to be extracted. Finally, some corrections are applied to reduce the noise and eliminate false positives near the brain ventricles. This process yields the lesion mask (Fig. 1h).

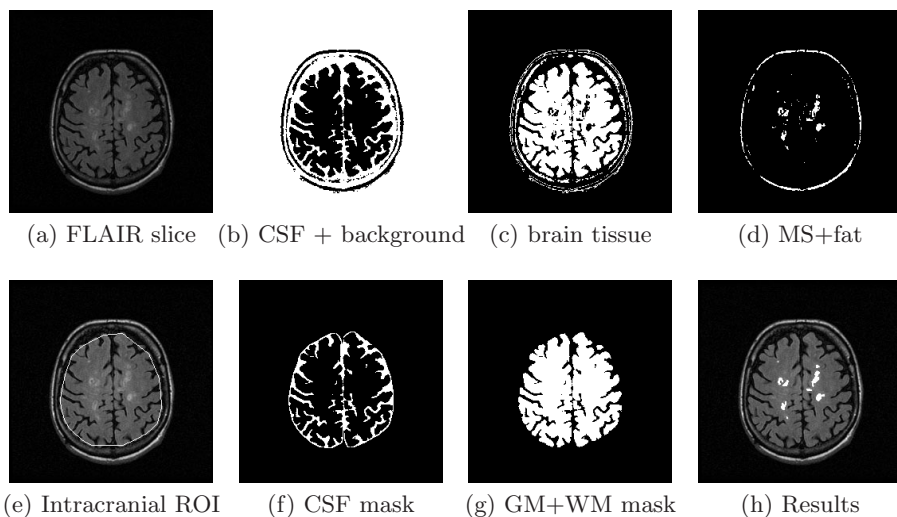


Fig. 1. Initial segmentation steps

Although the main goal of the fast method is to produce the lesion (demyelination plaques) mask, white and grey matter mask, as well as the CSF mask are also generated. The masks can be used to estimate the object and background parameters and choose seed points for the fuzzy-connectedness-based segmentation method.

4 Automated Fuzzy Connectedness Segmentation

The methodology proposed for the ‘fast’ segmentation may yield inaccurate results in a presence of local signal inhomogeneities for selected lesions. The inhomogeneities may be caused by: MR partial volume effects (likely to occur), acquisition noise (unlikely) or insufficient compensation of MR bias field during acquisition or preprocessing.

To increase the robustness of the segmentation, a new relative fuzzy connectedness-based processing step has been designed. The main goal of the step is to increase the processing accuracy by reducing the undersegmentation effect of the ‘fast’ method. First, the results of ‘fast’ segmentation are used in order to determine a fuzzy affinity parameters. Then, regions of analysis are determined for each 2D MR slice, based on already detected MS plaques and morphological thickening operation. Within each region, pixels labelled previously as demyelinated as well as not-demyelinated tissue are usually present. Finally seed points for the lesion and background are selected separately for each region of analysis, fuzzy connectedness is evaluated and the segmentation is performed.

4.1 Fuzzy Affinity Parameters Estimation

For the purpose of the MS lesions segmentation, a fuzzy affinity function form presented in Eq. 2 has been chosen with (Eq. 3) $G_1 \equiv g_1$, $G_2 \equiv g_2$ and weight $w = 0.5$.

Affinity parameters $m1$, $s1$, $m2$, $s2$ correspond to the average and standard deviation of intensity and local gradient, respectively. During the segmentation, they are estimated globally (within the entire MR volume). As the relative fuzzy connectedness approach is employed, two separate parameter sets have to be collected: one for MS lesions, and one for the ‘background’.

In order to estimate the demyelinated tissue parameters, first a morphological thinning of previously obtained lesion mask is performed. A mean and standard deviation of intensity distribution within the regions of MR series, that match the thinned mask (Fig. 2a-I) are used as $m1$ and $s1$. To estimate a local gradient, spatially averaged MR slices (3×3 averaging filter) are subtracted from the original slices. Parameters $m2$ and $s2$ are then selected as a mean and standard deviation of the obtained series (where only the regions matching the thinned mask are considered).

Similar method is used to select parameters of the ‘background’, i.e. the parameters of white and grey matter surrounding the lesions. First a morphological thickening of the lesion mask is performed. The intersection of the thickened

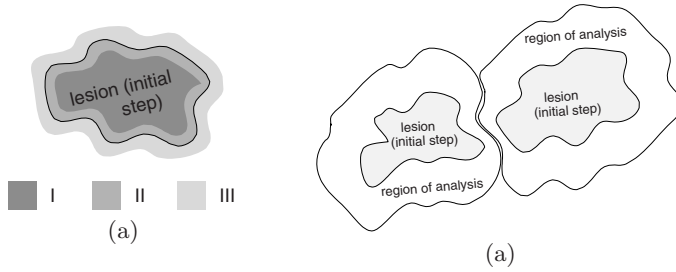


Fig. 2. Fuzzy affinity parameters estimation of the demyelinated tissue – region I, ignored – region II, background – region III (a) Schematic view of region of analysis (b)

mask and white and grey matter mask of ‘fast’ processing step is used to select pixels for parameters estimation (Fig. 2a-III). The m_1 , s_1 , m_2 , s_2 are estimated as for the demyelinated tissue.

4.2 Selecting Regions of Analysis and Segmentation Seed Points

To evaluate the fuzzy connectedness for a given image with respect to a selected seed point, all possible paths connecting the seed point and each image pixel have to be analysed. For a standard MR slice (e.g. 256×256 or 512×512), this is usually a very resource-consuming operation. Should a wider neighbourhood window mask be used (5×5 or 7×7), the processing time alone can increase beyond an acceptable level. As the main goal of the fuzzy connectedness processing is to increase the accuracy of segmentation for already detected lesions, it is not necessary to analyse a full data set at this step. Implementation of the 5×5 neighbourhood window within a region, containing a single MS plaque and some amount of surrounding tissue, permits the most typical cases of undersegmentation to be eliminated without much impact on the overall analysis time. In order to separate the MR slice into analysis regions, the demyelination lesion mask of the ‘fast’ segmentation step is used. First, a morphological labelling is performed. Then, for all the labelled regions simultaneously, morphological thickening is executed. This operation preserves the total number of regions with respect to a given neighbourhood window, and yields a region of analysis mask (Fig. 2b).

In each obtained region, a separate seed points of relative fuzzy connectedness is selected. The global s_1 , m_1 , s_2 , m_2 parameters are used to evaluate the region pixels against an index function:

$$mf(\mathbf{c}) = 0.5 \exp\left(-\frac{1}{2s_1^2}(f(\mathbf{c}) - m_1)^2\right) + 0.5 \exp\left(-\frac{1}{2s_2^2}(|\bar{f}(\mathbf{c})| - m_2)^2\right), \quad (7)$$

where $f(\mathbf{c})$ denotes the pixel signal level and $\bar{f}(\mathbf{c})$ represents the pixel of spatially averaged MR slice (cf. 4.1).

Pixels with the highest index function values are selected as seed points for the segmentation (Fig. 3a).

4.3 Segmentation

Selection of the seed points (Fig. 3a) permits the segmentation to be performed for each analysis region. First, the fuzzy affinity is found using the global parameters (4.1). Then, the fuzzy connectedness is evaluated with respect to the demyelinated tissue and background seed points (Fig. 3b-c). Finally, the objects are extracted (Fig. 3d).

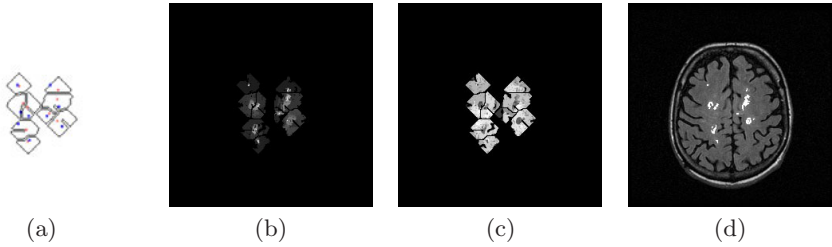


Fig. 3. Demyelinated tissue - diamonds - and background seed points - squares (a), fuzzy connectedness w.r.t. demyelinated tissue seed points (b), w.r.t. background seed points (c), and final segmentation result (d)

5 Results and Conclusions

The automated fuzzy connectedness segmentation method has been implemented in the Matlab software environment, and - with the ‘fast’ segmentation method - has been used in a Multiple Sclerosis Computer Aided Diagnosis (CAD) workstation.

In this section, results of an interobserver study have been presented. The automated segmentation has been tested against a reference set, containing results of radiological examination. During the verification, first ‘fast’ and fuzzy connectedness segmentation steps (denoted here as I and II) have been evaluated separately. Then, the combined (logical ORing) demyelination mask (I+II) have been examined.

For the evaluation purposes, three indexes have been calculated to compare the reference and automated results in terms of True Positives, True Negatives, False Positives and False Negatives: sensitivity

$$P = \frac{TN}{TN + FP} = 1 - \frac{FP}{TN + FP}, \quad (8)$$

specificity

$$P = \frac{TN}{TN + FP} = 1 - \frac{FP}{TN + FP}, \quad (9)$$

and and kappa statistics-based [10, 11] Dice Similarity Coefficient

$$DSC = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}. \quad (10)$$

Table 1. Processing results

no.	Vol. (ml)	Sensitivity (%)			Specificity (%)			Similarity (%)		
		I	II	I+II	I	II	III	I	II	III
1	0.88	92.21	87.34	94.48	99.992	99.993	99.989	68.02	67.50	63.06
2	1.64	71.84	58.53	76.62	99.998	99.998	99.996	79.36	68.53	78.70
3	2.91	60.31	60.41	68.96	99.998	99.996	99.994	72.19	69.61	74.21
4	3.33	89.36	79.74	91.16	99.996	99.998	99.995	89.47	85.74	89.02
5	7.51	92.95	86.62	93.83	99.954	99.983	99.951	75.58	84.05	75.04
6	10.7	92.03	83.39	93.55	99.924	99.966	99.916	72.63	79.20	71.66
7	12.3	91.74	80.15	93.16	99.936	99.969	99.93	77.59	79.31	76.94
8 ₁	27.9	94.39	83.72	95.10	99.858	99.933	99.849	79.31	81.96	78.74
9 ₂	31.0	87.40	70.62	90.47	99.960	99.980	99.947	88.02	80.16	88.24
8 ₁	32.9	94.87	82.44	96.20	99.915	99.977	99.908	87.39	87.48	87.41
9 ₂	34.6	83.38	62.27	85.35	99.978	99.974	99.962	88.27	73.82	87.66
10	0.26	28.26	17.39	29.35	99.998	99.998	99.997	28.11	20.51	26.87

The reference set has included 10 cases (460 images) randomly chosen from 21 available cases of confirmed Multiple Sclerosis. FLAIR MR images in reference set have been segmented by an experienced radiologist using tools implemented in the CAD workstation (Tab. 1). To estimate the impact of verification methodology, two cases have been additionally chosen to perform intraobserver study² using alternative supervised segmentation approach ($DSC_8 = 84.04\%$, $DSC_9 = 83.12\%$).

The introduced fuzzy connectedness-based segmentation stage (II) yields a higher specificity (Tab. 1) yet smaller sensitivity than the ‘fast’ segmentation stage (I). The highest sensitivity for combined (I+II) segmentation indicates, that the results of both stages are complementary. The DSC index shows a high correlation between the reference and evaluated results. During an examination of the worst case (10), several weak points of presented methodology have been identified. The small lesions close to the CSF have been labelled false positives, and small lesions close to the skull have been detected as false positives. With total of 96 pixels labelled as presenting demyelinated tissue in reference results, this accounts on the poor performance of the ‘fast’ segmentation.

References

1. Alfano, B., Brunetti, A., Larobina, M., et al.: Automated segmentation and measurement of global white matter lesion volume in patients with multiple sclerosis. *J. Magn. Reson. Imaging* 12, 799–807 (2000)
2. Udupa, J.K., Wei, L., et al.: Multiple sclerosis lesion quantification using fuzzy-connectedness principles. *IEEE Trans. Med. Imag.* 16, 598–609 (1997)

² Verification against both intraobserver results has been included in table (Tab. 1) as cases 8₁/8₂ and 9₁/9₂.

3. Kikinis, R., Guttman, C.R., Metacalf, D., et al.: Quantitative follow-up of patients with multiple sclerosis using MRI. Technical aspects *Journal of Magnetic Resonance Imaging*, 519–530 (1999)
4. Sajja, B.R., Datta, S., He, R., Narayana, P.A.: A unified approach for lesion segmentation on MRI of multiple sclerosis. In: *Proceedings of the 26th Annual International Conference of the EMBS*, vol. 1, pp. 1778–1781 (2004)
5. Kawa, J., Pietka, E.: Kernelized fuzzy c-means method in fast segmentation of demyelination plaques in multiple sclerosis. In: *Proceedings of the 29th Annual International Conference of the EMBS*. IEEE Computer Society Press, Los Alamitos (2007)
6. Udupa, J.K., Samarasekera, S.: Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation *Graph. Models Image Processing* 58, 246–261 (1996)
7. Pednekar, A.S., Kakadiaris, I.A.: Image segmentation based on fuzzy connectedness using dynamic weights. *IEEE Trans. Image Processing* 15, 1555–1562 (2006)
8. Kawa, J., Pietka, E.: Kernelized Fuzzy c-means method in segmentation of demyelination plaques in multiple sclerosis. In: *Int. X.I. (ed.) XI Int. Conference on Medical Informatics and Technologies - MIT 2006*, Skalmierski, Gliwice, pp. 20–27 (2006)
9. Bezdek, J.C.: *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York (1981)
10. Dice, L.: Measures of the amount of ecological association between species. *J. Ecology* 26, 297–302 (1945)
11. Zijdenbos, A.P., Dawant, B.M., Margolin, R.A., Palmer, A.C.: Morphometric analysis of white matter lesions in MR images: method and validation. *IEEE Trans. Med. Imag.* 13, 716–724 (1994)

Computer-Interactive Methods of Brain Cortical Evaluation

Anna Czarnecka¹, Marek J. Sasiadek¹, Elzbieta Hudyma²,
Halina Kwasnicka², and Mariusz Paradowski²

¹ Department of General Radiology, Interventional Radiology and Neuroradiology,
Chair of Radiology, Wrocław Medical University

² Institute of Applied Informatics, Wrocław University of Technology

Summary. The subject of the paper is the evaluation of a brain atrophy depending on computed tomography (CT) images. There are two computer aided methods under investigation. The first one is a semi-automatic volumetric method, the second is a fully automatic one and depends on the fractal dimension. Results of the experiment with 68 patients show that both methods are able to estimate the brain atrophy very similarly to the evaluation made by expert radiologist with the coefficient of correlation in the range of 0.7-0.9. The great advantage of the automatic method comparing to volumetric method is generating results in a very short time, that counts in ordinary clinical practice and could be used to experiment with huge patient groups.

1 Introduction

Atrophy of the brain is associated with physiological aging, but also with degenerative diseases of the central nervous system, including dementive disorders. The increasing average age of the societies in developed countries results in increased frequency of dementive diseases, which becomes more and more serious social and medical problem. Therefore there are efforts to improve effectiveness of diagnosing dementia as early as possible, and thus enable early treatment [6, 8, 5].

Imaging methods, including computed tomography (CT) play an important role in dementia diagnostics. The aim of CT is to rule out the reversible causes of dementia (e.g. brain tumor, normal pressure hydrocephalus, infection) and to assess brain atrophy. It is essential to define if the brain atrophy detected in the patient is caused by normal aging or dementive disease or overlapping of both processes. Therefore attempts are made to establish precise and reliable quantitative methods, which could answer this question [2, 3, 1, 4].

The aim of the study is evaluation of the authors' methods of measurement of cortical and subcortical atrophy based on calculation of the cerebrospinal fluid (CSF) spaces (outside the brain, it is consistent with cortical atrophy and intraventricularly, which is consistent with subcortical atrophy).

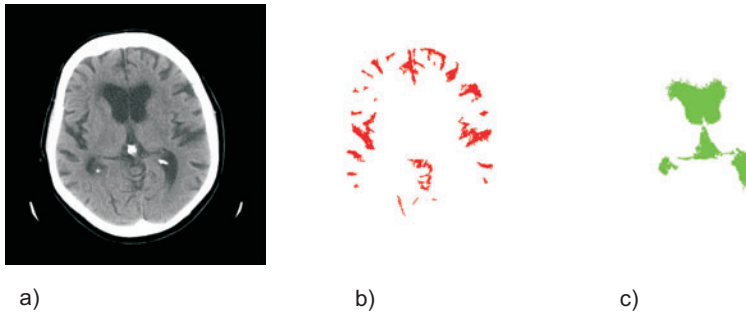


Fig. 1. The example of CT slice image and its processed images: a) the original CT, b) the cortical brain atrophy, c) the subcortical brain atrophy

In daily practice the brain atrophy is assessed visually by radiologist, which is characterized by:

- Simple and quick evaluation of the atrophy,
- Non-objectivity of a measurement, dependent on the experience and ability of an expert.

The aim of the study is to evaluate the usefulness of the proposed methods of the analysis of CT images using computer processing and artificial intelligence approaches (Fig. 1). The possible advantages of such computer system are:

- objectivity and repeatability of measurements,
- short time of assessment (directly after performing CT),
- low cost of assessment.

In this paper two computer methods are presented. The obtained results are compared with the evaluation of the same images made by an experienced radiologist.

2 Methods of Cortical and Subcortical Atrophy Evaluation

Two methods of the brain atrophy computer evaluation are developed:

1. The semi-automatic method, using volumetric software of CT unit, corrected by the radiologist (volumetric method).
2. The fully automatic method, based on the authors segmentation idea and fractal geometry (fractal method).

2.1 The Volumetric Method

In the semi-automatic method source image of the brain are transferred to the workstation of CT unit. The radiologist selects range of density in Hounsfield

units (HU) for the regions of interest. The authors choose the range from -5 to $+20$ HU. In the next stage the radiologist corrects manually the areas found by computer by outlining anatomic structures borders. Correction has to be performed for each section, separately for extracerebral areas of CSF (cortical atrophy) and intraventricular space (subcortical atrophy). In the last stage, after confirming the labeled regions of interest, the volume of both spaces in cm^3 is calculated by volumetric program.

2.2 The Fractal Method

The method is fully automatic and operates in two stages: In the first one each image of a CT axial slice is treated as a bitmap in the gray scale. A dedicated segmentation algorithm is employed to find the cerebrospinal fluid area. The algorithm is based on pixels hue and placement in the original image. More details on the algorithm can be found in [9]. After the segmentation stage a binary image representing the cerebrospinal fluid is available. The goal of the second stage is to find the best measure of this object shape.

The brain, like other organisms and real-world phenomena, exhibits fractal properties. It can be useful to calculate the fractal dimension of a set of sampled data. The fractal dimension measures cannot be calculated exactly but their values have to be estimated. The fractal characteristic is a real number and defines the level of the object complexity.

Casual understanding of a set dimension is the number of independent parameters required for the localization of a point within this set. The topological dimension of a set is the mathematical concept which models this idea. For example, a point in the plane is described by two independent parameters (the Cartesian coordinates of the point), so in this sense, the plane is two-dimensional. Of course, topological dimension is always a natural number. However, topological dimension is quite useless on certain highly irregular sets such as fractals.

There are various closely related notions of possible fractional dimensions [7, 10]. These notions (topological dimension, Hausdorff dimension, Minkowski-Bouligand dimension, information dimension, correlation dimension) give the same value for many shapes. But sometimes they give different values for some highly irregular curves, particular for multifractals. The most popular and universal method is box-counting dimension known as Minkowski-Bouligand dimension. It is an easy method for fractal dimension estimation, also useful for nonfractal objects.

The space is divided into a grid of boxes of size S . Afterward, a number of boxes scale $N(S)$ that would contain part of the attractor is counted. The process is repeated for the decreasing size of grid. The formula of the box-counting dimension is the following:

$$D = \frac{\log(N(S_{n+1})) - \log(N(S_n))}{\log(1/S_{n+1}) - \log(1/S_n)}, \quad (1)$$

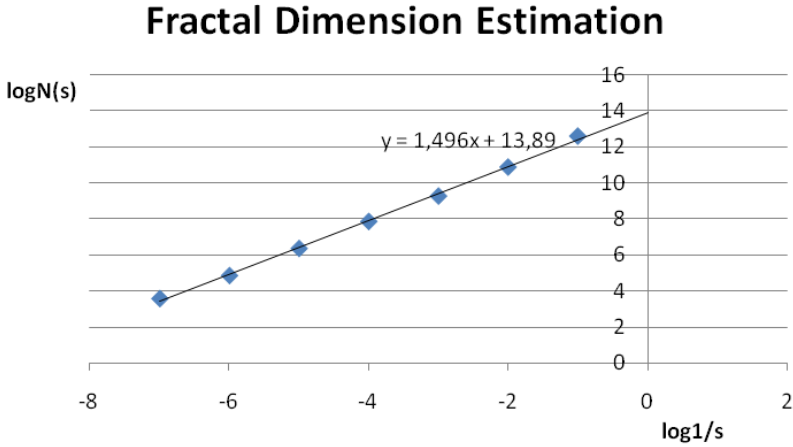


Fig. 2. The curve approximated points $\log N(s)$ vs $\log 1/s$. The slope of the line is the fractal dimension $D = 1.496$.

where $N(S)$ – the number of boxes which contain fractal for grid S ,
 S_{n+1} – the next grid, $S_{n+1} > S_n$.

An example of box-counting dimension estimation is presented in Fig. 2.

3 Research Description

The study is performed on a group of 68 patients for whom clinical examination and neuropsychological tests confirmed cognitive impairment, compatible with dementia. For all patients CT with two-row spiral unit is performed (Dual HiSpeed / GE Medical Systems). Five to 17 axial slices are obtained for each patient.

CT images are firstly assessed visually by radiologist to evaluate roughly cortical and subcortical atrophy and to rule out other lesions, which could explain patient's symptoms. Visual assessment is based on source images analysis, with classification of the atrophy in 0-3 scale, where: 0 means no atrophy (normal appearance for age), 1 – slight atrophy, 2 – moderate atrophy and 3 – severe atrophy.

The automatic atrophy evaluation on the basis of CT images requires the suitable graphic preprocessing of each axial image. It results in a binary image of intracranial areas, which are the subject of our interest, i.e. cortical and subcortical atrophy.

The first method (volumetric method) requires the expert radiologist knowledge in the each stage of execution:

- Preliminary definition of density range in HU to label the region of interest using the volumetric software.

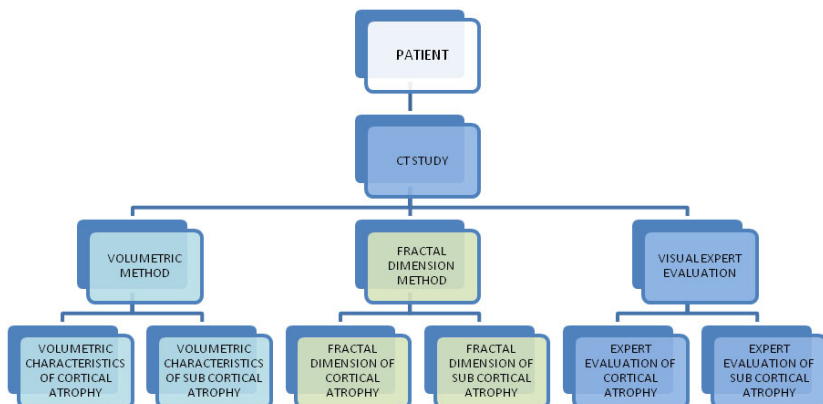


Fig. 3. The schematic diagram of the research

- Visual assessment of the areas compatibility labeled by the computer program with the anatomical borders in CT study.
- Manual correction of the borders labeled by the program by outlining structures or widening the density range for each section to fit the labeled areas to the anatomical structures.
- Calculation of labeled areas volumes by the computer program (separately for subcortical and cortical atrophy).

The second method (the fractal method) does not need a human interference and relies on automatic processing:

- Preprocessing of the CT images using the authors cerebrospinal fluid segmentation algorithm [9].
- Fractal dimension calculation of objects found in the previous stage. This is the box-counting dimension presented in the previous section.
- Fractal dimension mean value calculation for all axial slices of the patient CT study.

The diagram in Fig. 3 shows the schema of the research.

4 Research Results

As a result in our experiment we receive a set of data for each patient. Atrophy evaluation comparison is done separately for cortical and subcortical atrophy regions. The aim of the comparison is to check the proposed methods effectiveness and to compare them with visual assessment performed by radiologist.

Fractal dimension calculated for interesting areas is a real number in the range 1.1 - 1.8. Volumetric values are real numbers in the range 0.7 - 20.3. These values concern the regions of interest containing CSF, referred to the volume

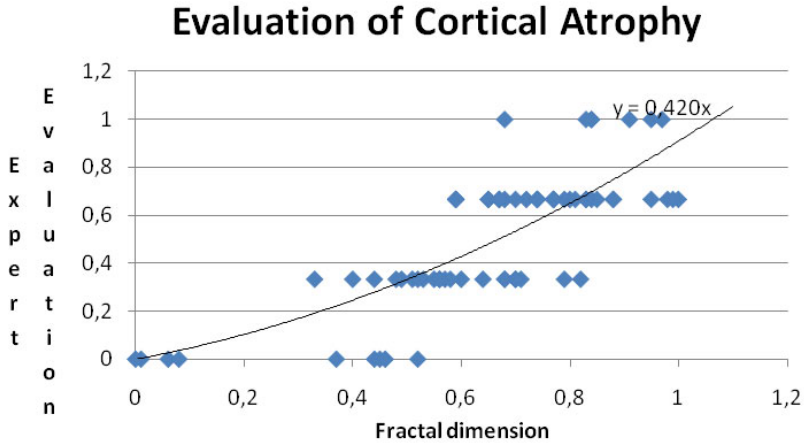


Fig. 4. Relationship between expert evaluation of cerebral atrophy and its evaluation by fractal dimension

of the whole intracranial space. To facilitate the comparison all the values are normalized in the range 0-1.

We check an assumption that measurement of cortical and subcortical atrophy by computer methods is similar to visual evaluation atrophy by the radiologist. Each pair presented in Fig. 4 specifies Cartesian point coordinates in the plane. The pattern of points allows to estimate the function between values. The function is calculated by the least square method and shown on the plot. The plot shows the relationship between the expert visual evaluation and the fractal dimension evaluation of cortical atrophy.

A measure of force between the two elements in the couple is the correlation coefficient ρ . It is estimated from a sample population r and is determined as Pearson coefficient. Correlation coefficients for the couples of characteristics are presented in Table 1. Results rely on the research for 68 patients. The two numbers in the table indicate the cortical and subcortical atrophy, respectively. On the significance level of 0.05, sample size of 68, the critical value $r_{0.05;66}$ is equal to 0.250. Because $|r| > r_{\alpha,v}$ there is significant statistic relation between evaluation of cortical and subcortical atrophy made by the expert and by the

Table 1. Correlation coefficients for the brain atrophy (cortical/subcortical) between the couple of values

	Visual expert evaluation	Fractal dimension	Volumetric coefficient
Visual expert evaluation	1	0.778/0.749	0.815/0.801
Fractal dimension	0.778/0.749	1	0.838/0.772
Volumetric coefficient	0.815/0.801	0.838/0.772	1

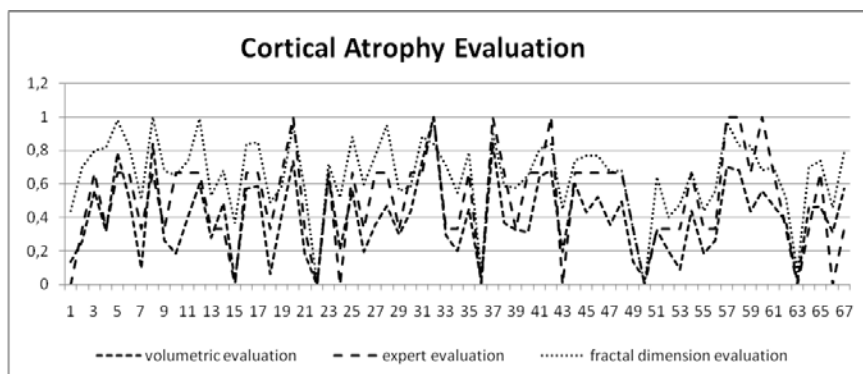


Fig. 5. Cortical atrophy normalized evaluation comparison for three methods and all checked patients

computer methods. In all performed experiments there is a strong relation, in range 0.7-0.9. The plot in Fig. 5 presents evaluations of the cortical atrophy for all 68 patients calculated by all three methods. Values are normalized in the range 0-1.

5 Concluding Remarks

Semi-automatic method is very compatible with visual evaluation made by the radiologist. However, it is very time-consuming because an operator needs round about 2.5 hours to prepare results for one patient. The preprocessing stage results are strongly dependent on the operator and his anatomic knowledge. This makes the results non repetitious under similar conditions. Expert's evaluation has the same flaw.

Fully automatic method is a little less compatible with the expert evaluation than the previous one. However, its largest advantage is the independence from an operator, repeatability, quickness and low cost.

These advantages of the automatic method enable us to save time and reduce cost in ordinary clinical practice. Additionally, research for a norm of the brain atrophy associated with physiological aging is made possible using huge patient groups. Results of such research improve early diagnosing dementia and thus enable early treatment.

Repeatability of the method enables a longitudinal study with serial scans on subjects for detecting progressive cerebral atrophy. It allows the determination of differences between rates of atrophy in patients with dementive disorders and treatment efficacy monitoring. Additionally, good results of the automatic approach encourage us to extend the method with more complex techniques of fractal geometry and image processing.

Acknowledgement. This work is financed from the Ministry of Science and Higher Education resources in 2006 - 2008 years as a research project N518 022 31/1338.

References

1. Chetelat, G., Baron, J.C.: Early diagnosis of Alzheimer's disease: contribution of structural neuroimaging. *Neuroimage* 18(2), 524–525 (2003)
2. Fox, N.C., Schott, J.M.: Imaging cerebral atrophy: normal ageing to Alzheimer's disease. In: *Lancet* 2004, vol. 363, pp. 392–394 (2004)
3. Frisoni, G.B.: Structural imaging in clinical diagnosis of Alzheimer's disease: problems and tools. *J Neurol Neurosurg Psychiatry* 70, 711–718 (2001)
4. Golebiowski, M.: Dementive diseases (in Polish). In: Walecki, J. (ed.) *Advances in neuroradiology*, Warsaw. Polish Foundation for Science Advancement (PFUN), pp. 247–264 (in Polish) (2007)
5. Kotapka - Minc, S., Szczudlik, A.: Dementia In: *Diagnosis and Treatment of Dementia*, Czelej, Lublin, pp. 11–23 (2006)
6. Leszek, J. (ed.): *Dementive Diseases – theory and practice*, Continuo, Wroclaw (2003)
7. Peitgen, H.O., Jurgens, H., Saupe, D.: *Chaos and Fractals: New Frontiers of Science*, parts I and II. Warsaw, PWN (in Polish) (1995)
8. Ritchie, K., Lovestone, S.: The dementias. In: *Lancet* 2002, vol. 360, pp. 1759–1766 (2002)
9. Tabakov, M., Kwasnicka, H., Krynicki, K.: A Rule-based Region Growing Fuzzy Segmentation System for Pathological Brain Computed Tomography Images (in preparation) (2008)
10. Walecki, P., Trabka, J.: Fractal measures in neuroimaging. *Artificial Intelligence in Biomedical Engineering* (in Polish) (2004)

Automatic Registration of MRI Brain

Piotr Zarychta¹ and Anna Zarychta-Bargieła²

¹ Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice
piotr.zarychta@polsl.pl

² Hospital in Dąbrowa Górnicza, ul. Szpitalna 13, 41-300 Dąbrowa Górnicza

Summary. This paper shows an automatic registration method of the A- and B-group of MRI brain for the same patient. The time difference between recording the A-group and the B-group is equal six months. The automatic registration of the A- and B-group of MRI brain is based on the entropy and energy measures of fuzziness. First, two sequences (A- and B-group) are converted to a fuzzy representation. Then, the entropy and energy measures are employed in the NCC and GD methods. The alignment based on energy and entropy fuzzy measures shows a significant improvement in comparison with the implementation of the original image.

1 Introduction

Medical image registration is a difficult and complex problem. Clinical diagnosis and treatment verification is often supported by imaging modalities, providing functional and anatomical information. The anatomical images provide information of the anatomic structure of the human body. In many instances it is necessary to integrate the information obtained from two or more studies of the same patient. But the differences in patient positioning and different image acquisition parameters require the registration of these images before overlapping them. In medical application, registration process should be automatic, reliable and easy to use [7].

In this paper a registration method, based on fuzzy image idea is discussed. Since the analysis is performed on one MRI study, a registration procedure has to precede the ROI extraction. Due to the time difference equals six months, the brain can be a bit shifted between acquisition both series (fig. 1).

2 A Fuzzy Image Idea

The idea of a fuzzy image is based on the concept of a fuzzy signal, introduced by Czogala and Leski in [3, 6]. Let us consider an image $X(N, M)$ at the size of $N \times M$ with pixels $I(n, m)$ and $n = 1, 2, \dots, N$; $m = 1, 2, \dots, M$. Moreover, the image is scanned within a window at the size of $(2k + 1) \times (2k + 1)$. The idea of a fuzzy image derived from the original image is based on two assumptions: if no fuzzy uncertainty is considered in the image, a fuzzy image pixel

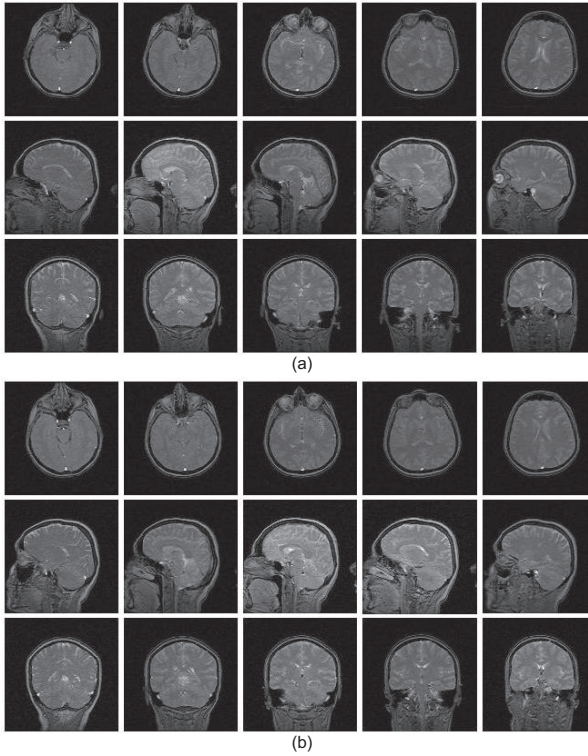


Fig. 1. MRI brain - signal from data base clinical hospital (a) A-group (b) B-group

$I(n, m, k)$ is reduced to a real number $I(n, m)$, which is referred to as a singleton. The measure of fuzziness is equal to zero and the information increases if the original image changes, as does the dynamic, as smaller amount of information

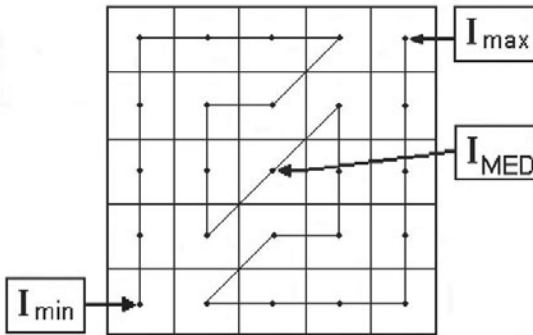


Fig. 2. Sorted pixels within the window

is conveyed by the image. In order to construct a fuzzy image from crisp image, the pixels within the window are sorted into an increasing order is shown in fig. 2. The median value is located at the (n, m) position within the window, ie. $I_{MED}(n, m, k) = I(n, m, k)$. Based on the median operation performed within windows scanned over the image, a membership function μ is defined. First we assume, that $\mu_{n,m,k}(I_{min}(n, m, k)) = 0$, $\mu_{n,m,k}(I_{max}(n, m, k)) = 0$ and $\mu_{n,m,k}(I_{MED}(n, m, k)) = 1$. The remaining elements are defined according to the following formula

$$\mu_{n,m,k}(I_{i,j}(n, m, k)) = \mu = \frac{(2k+1)^2 - (2d_{i,j} + 1)^2 + 1}{(2k+1)^2}, \quad (1)$$

where $d_{i,j} = \max(i, j)$ for $i, j = 0, \dots, k$. In order to discriminate against a certain level of membership $\lambda \in [0, 1]$, the Heaviside pseudofunction $\mathbf{1}(\cdot)$ is introduced. Implementation of the Heaviside function is defined as

$$\mu^\lambda = \mu \mathbf{1}(\mu - \lambda). \quad (2)$$

The introduction of the λ level into the measurement of fuzziness allows the insignificant membership degree to be removed. A concept of the entropy measure of fuzziness has been introduced in [4] and implemented to biomedical signals in [3, 9, 10, 11]. The entropy measure of fuzziness is a mapping from the set of all fuzzy subsets of a base set X into the nonnegative real numbers. It can be expressed as

$$H(A, \lambda) = \int_X h_\lambda(A(x)) d\nu, \quad (3)$$

where $A : X \rightarrow [0, 1]$ is any ν -measurable function, $d\nu = dx$ or $d\nu = p(x)dx$, where $p(x)$ stands for a probability density function; $h : [0, 1] \rightarrow R_+$ is an increasing function in $[0, 0.5]$, a decreasing function in $[0.5, 1]$, and $h(0) = h(1) = 0$. Substituting (2) into (3) yields

$$H(\mu^\lambda) = F_1 \left(\sum_{i=1}^{2k} \sum_{j=1}^{2k} h_\lambda(\mu^\lambda p(I_{i,j}) \Delta I_{i,j}) \right), \quad (4)$$

where $F_1 : R_+ \rightarrow R_+$ is an increasing function and $F_1(z) = 0$ if $z = 0$, $I_{i,j}$ denotes $I_{i,j}(n, m, k)$, and $h_\lambda(z)$ is defined as

$$h_\lambda(z) = \begin{cases} h(z) & \text{if } z \in [\lambda, 1 - \lambda] \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

and $\Delta I_{i,j}$ is a gradient between neighboring pixel values as marked in fig. 2, and the probability density function $p(I_{i,j})$ is obtained from a histogram, by dividing the number of pixels by $2k + 1$.

The membership functions, as defined above, serves also as a basis for the energy extraction. The energy measure is expressed as

$$E(A, \lambda) = \int_X f_\lambda(A(x)) d\nu, \quad (6)$$

where

$$f_\lambda(z) = \begin{cases} f(z) & \text{dla } z \in [\lambda, 1] \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

The final formula for the energy measure can be written as

$$E(\mu^\lambda) = F_2\left(\sum_{i=1}^{2k} \sum_{j=1}^{2k} f_\lambda(\mu^\lambda p(I_{i,j}) \Delta I_{i,j})\right), \quad (8)$$

where $F_2 : R_+ \rightarrow R_+$ is an increasing function and $F_2(z) = 0$ if $z = 0$.

3 Similarity Measures

Once fuzzy images are generated similarity measures are to be employed. Here are two main groups of similarity measures: feature-based and intensity-based. Feature-based measures first process the images in order to obtain significant information, which can be used to judge the similarity. This can be the position of significant landmarks, or the quantitative description of a particular anatomical structure obtained by segmentation [1]. Intensity-based measures use a full raw image information. A similarity measure is derived using intensity values in registered images. However, one may consider introducing the region of interest in order to omit non relevant image parts. Working with this type of measure is often referred to as voxel property based registration. The main advantage is that registration can be executed right after the image acquisition and definition of an initial pose [1, 2]. Intensity-based measures compare the intensity values of both images pair-wise at the same pixel positions. Subsequently one single value is composed out of it with a certain scheme. An advantage of this type of measure is that it can be used not only with 2D images, but with data of arbitrary dimensions, as no spatial information is considered. In our study three measures have been introduced: sum of squared difference (SSD), sum of absolute difference (SAD) and normalized cross correlation (NCC)

$$NCC = \frac{\sum_{n,m \in T} [I_1 - \bar{I}_1] [I_2 - \bar{I}_2]}{\sqrt{\sum_{n,m \in T} [I_1 - \bar{I}_1]^2} \sqrt{\sum_{n,m \in T} [I_2 - \bar{I}_2]^2}}, \quad (9)$$

where $I_1 = I_1(n, m)$ and $I_2 = I_2(n, m)$, T is the overlap domain of the images, N is the number of pixels in T , \bar{I}_1 and \bar{I}_2 denote the mean intensity values in I_1 and I_2 images, respectively.

This means that neither contrast nor brightness affect the similarity measure [8]. By using horizontal H and vertical V Sobel templates [5] four gradient

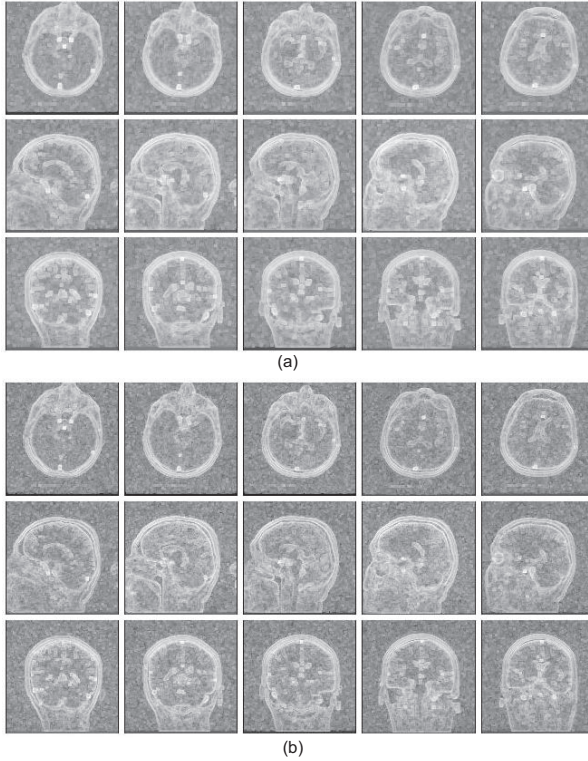


Fig. 3. A-group MRI brain (a) entropy and (b) energy measure of fuzziness

images $\frac{dI_1}{dn}$, $\frac{dI_2}{dn}$, $\frac{dI_1}{dm}$ and $\frac{dI_2}{dm}$ are created. Then, NCC of the horizontal and vertical gradient images are calculated, and are referred to as NCC_{difH} and NCC_{difV} , respectively. The final value NCC_{dif} is the average of both measures. Gradient difference (GD), pattern intensity (PI), gradient correlation (GC), and sum of local normalized correlation ($SLNC$) are included in the group of measures based on spatial information. A very important advantage of the gradient measure is the ability to remove low spatial frequency differences between two images (e.g. caused by soft-tissue structure) allowing the similarity measure to be based on the edge information [2, 8]. The gradient difference (GD) measure is defined as

$$GD = \sum_{n,m \in T} \frac{A_H}{A_H + [I_{difH}]^2} + \sum_{n,m \in T} \frac{A_V}{A_V + [I_{difV}]^2}, \quad (10)$$

where A_H and A_V are constants, which denote variance of the reference images, has been tested. It evaluates two gradient images I_{difH} and I_{difV} , where I_{difH} and I_{difV} denote $I_{difH}(n, m)$ and $I_{difV}(n, m)$, respectively.

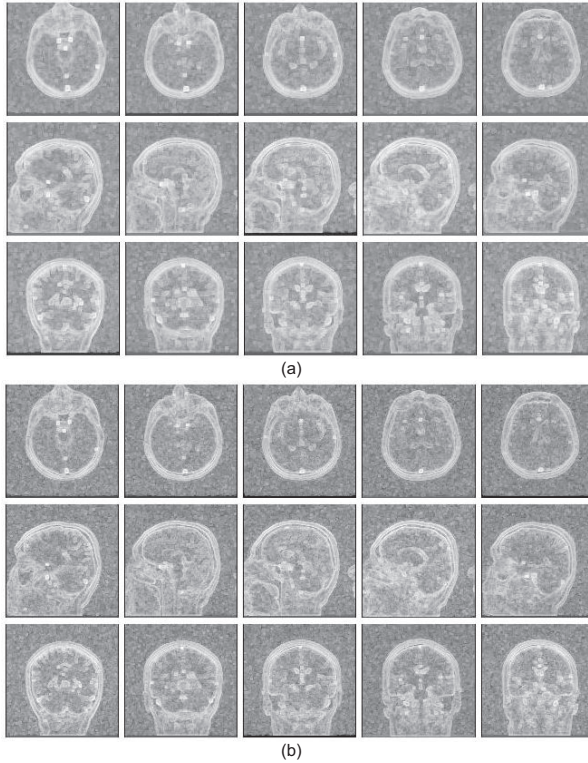


Fig. 4. B-group MRI brain (a) entropy and (b) energy measure of fuzziness

4 A- and B-Group Images Registration

The registration process relies on a comparison of each A-group MRI brain slice with each B-group MRI brain slice. The bigger the similarity measure the stronger the slice similarity. So for each A-group MRI brain slice is finding only one B-group MRI brain slice, for which the similarity measure is the biggest. In fig. 5 and fig. 6 are shown results of registration process for three cases. For slice no. 5 and no. 11 of A-group the similarity measure is the biggest for the same slice number of B-group (fig. 5), but for slice no. 7 of A-group the similarity measure is the biggest for slice no. 8 of B-group. On the basis of results analysis for similarity measure we can certify, that slices no. 6 to no. 10 of A- and B-group are shifted in relation to themselves.

In presented cases all measures give correct results, but measures for original images (GD , NCC , NCC_{dif}) are sensitive to noise and interference. Two measures (for fuzzy images $NCC_{norm.H(A)}$ and $NCC_{norm.E(A)}$) show strong discriminative power and are implemented in the application software.

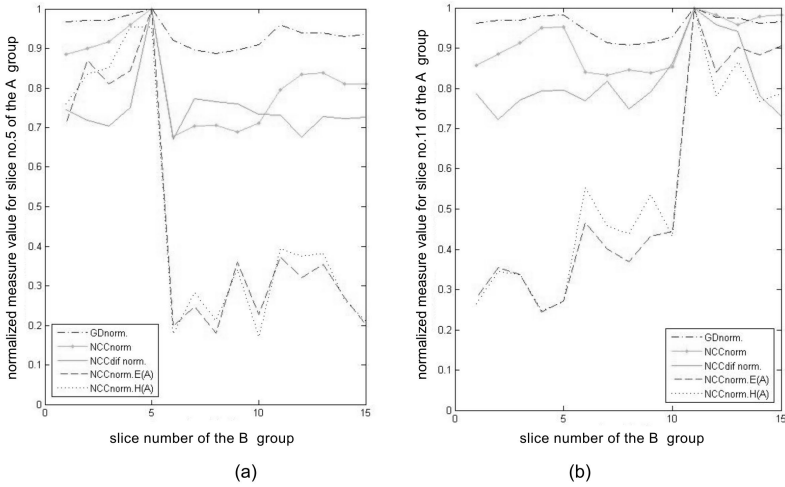


Fig. 5. Similarity measures for (a) slice no.5 and (b) slice no.11 of A-group

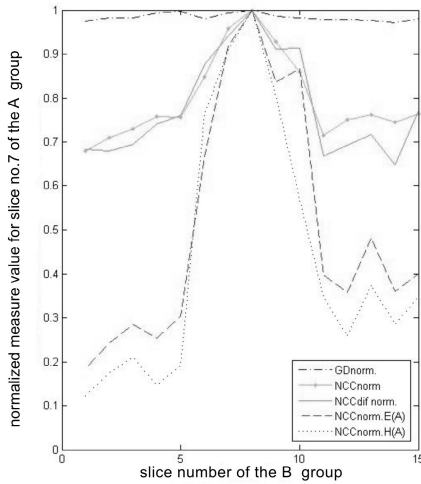


Fig. 6. Similarity measures for slice no. 7 of A-group

5 Results

At the registration face A- and B-group (fig. 1) are subjected to the analysis. In order to generate the fuzzy images (fig. 3 and fig. 4 the following parameters and functions have been selected: for energy measure of fuzziness $\lambda = 0$, $f(x) = x^2$, $F_2(z) = z^2$ and $d\nu = dx$, for entropy measure of fuzziness $\lambda = 0$, $h(x) = x^2$, $F_1(z) = z^2$ and $d\nu = dx$.

Calculation of energy and/or entropy measure of fuzziness for original images and next using normalized cross correlation, let for four times improving efficiency of recognizability corresponding slices from A- and B-group MRI brain series than by using normalized cross correlation for original slices.

Energy measure of fuzziness and entropy measure of fuzziness give quite similar results, but the energy measure of fuzziness is faster about 30%.

On the basis of the described methodology a software application was built. This application is dedicated for the cruciate ligament diagnosis. Registration process is first of three stages of that application and makes it possible to show PCL structure in 3D for T1- and T2-weighted MRI knee slices [11, 12].

Currently this methodology is testing for MRI brain. On the basis of preliminary investigation we can certify, that the automatic registration process has been performed correctly. The proposed method of broadening the idea of a fuzzy image creating and using the entropy and energy measure of fuzziness to the registration of MRI brain, seems to be very effective and promising.

References

1. Brown, L.: A survey of image registration techniques, *ACM Comp. Surveys* 24(4), 9–18 (1992)
2. Brown, L., Boulton, T.: Registration of planar film radiographs with computed tomography. In: *Proc. MMBIA*, pp. 42–51 (1996)
3. Czogała, E., Łęski, J.: Application of entropy and energy measures of fuzziness to processing of ECG signal, *Fuzzy sets and systems*, pp. 9–18 (1997)
4. Deluca, A., Termini, S.: A definition of non-probabilistic entropy in the setting of fuzzy set theory. *Information and Control* 20, 301–312 (1972)
5. Gonzalez, R., Woods, R.: *Digital Image Processing*. Prentice-Hall, Englewood Cliffs (2002)
6. Łęski, J.: A new possibilities of non-invasive electrocardiological diagnosis, *Scientific Reports of the Technical University of Silesia*, No 1233 (1995)
7. Nałęcz, M.: *Obrazowanie Biomedyczne. Biocybernetyka i Inżynieria Biomedyczna*. Exit, Warszawa (2003)
8. Penney, G., Weeseand, J., Little, J., Desmedt, P., Hill, D., Hawkes, D.: A Comparison of Similarity Measures for Use in 2D-3D Medical Image Registration. *IEEE Transactions on Medical Imaging* 17(4), 586–595 (1998)
9. Rozentryt, P., Czogała, E., Łęski, J.: Application of entropy and energy measures of fuzziness to heart rate variability analysis. *Medical Science Monitor* 2(5), 642–649 (1996)
10. Zarychta, P.: Cruciate Ligament Localization in T1- and T2-weighted MR Knee Images. In: *IFAC Workshop PDeS 2006*, pp. 203–207 (2006)
11. Zarychta, P.: Location and 3D visualization of the cruciate ligament in the MR knee images on the basis of fuzzy logic. PhD Thesis, Silesian University of Technology, Gliwice (2006)
12. Zarychta, P.: Posterior Cruciate Ligament - 3D Visualization. In: *V International Conference on Computer Recognition Systems, CORES*, pp. 695–702 (2007)

Magnetic Resonance Image Classification Using Fractal Analysis

Karol Kuczyński¹ and Paweł Mikołajczak²

¹ Maria Curie-Skłodowska University, Pl. M. Curie-Skłodowskiej 1, 20-031 Lublin, Poland

karol.kuczynski@umcs.lublin.pl

² Maria Curie-Skłodowska University, Pl. M. Curie-Skłodowskiej 1, 20-031 Lublin, Poland

mikfiz@goblin.umcs.lublin.pl

Summary. Fractal analysis is a reasonable choice in applications where natural objects are dealt with. Fractal dimension is an essential measure of fractal properties. Differential box-counting method was used for fractal dimension estimation of radiological brain images. It has been documented in this paper that this measure can be used for automatic classification of normal and pathological cases.

1 Introduction

Modern medical imaging techniques like MRI (Magnetic Resonance Imaging) are becoming more and more affordable and their quality is increasing continuously. Illnesses that are detected in this way include brain tumours and various abnormalities related to dementia (resulting from aging, Alzheimer's or Parkinson's disease) [1]. These illnesses are nowadays very common and are becoming a social issue. The number of images that are to be analysed by a radiologist is raising dramatically. However, there are no visible pathologies in most of acquired images. Relatively few of them need to be studied by a doctor thoroughly.

The authors' aim is to project, implement and test an algorithm that initially classifies MR images automatically. It is expected that fractal dimension can be used for this purpose.

It is known that images of many natural objects (mountains, coasts, trees, clouds, butterflies' wings [2], eye iris [3], metal corrosion [4], etc.) have fractal properties. Fractal properties of human brain cortex have also been carefully studied [5]. It happens to be self-similar in a way referred to as being a fractal, with a fractal dimension $D = 2.60$ [6] (the results vary and depend on calculation method). That is why fractal analysis is a reasonable choice in applications where natural objects are dealt with, including medical image processing and analysis.

Fractal has been defined by Mandelbrot [7] as a bounded set A in \mathbb{R} for which the Hausdorff-Besicovich dimension is strictly larger than the topological dimension. The concept of self-similarity is used to estimate the fractal dimension. If

A is the union of N_r non-overlapping copies of itself scaled down (or up) by a factor r , the fractal dimension is given by:

$$D = \frac{\log(N_r)}{\log(\frac{1}{r})}, \quad (1)$$

where $1 = N_r r^D$. It has been noticed, that fractal dimension correlates with human perception of surfaces (about 2 for smooth surfaces and close to 3 for rough surfaces) [8].

Equation 1 can be directly applied only to geometrical fractals. In image processing and analysis it has to be estimated phenomenologically. A survey of fractal dimension calculation methods can be found in [9]. Different variations of box-counting methods are the most popular. A binary image is placed on a grid of square blocks. The number of blocks N_r occupied by a part of the image is then calculated. The procedure is repeated for various grid sizes (r). It is expected that increasing the resolution of the grid, N_r should increase, too. The slope of linear regression of the $\log(N_r)$ versus $\log(1/r)$ is the fractal dimension estimation. If is stable along several orders of magnitude, it is expected to be accurate.

2 Materials and Methods

MRI T2-weighted axial 512×512 brain images were the subject of the analysis. Some of them include pathological changes related to dementia. The quality of MR images is still being improved. However, automated analysis is not a trivial task, due to some of their properties (high noise level, no standard intensity scale similar to Hounsfield's units for Computed Tomography, overlapping pixel intensities corresponding to different tissue, etc.).

The choice of fractal dimension calculation method is the essential issue. Typically, the main criterion for estimation of these methods is their accuracy, compared to the theoretical value of fractal dimension for a given geometrical object. In this task it is important to find a measure related to fractal features of an object to distinguish between normal and abnormal brain structures, rather than to calculate fractal dimension accurately.

Classical box-counting algorithms operate on binary images, so radiological grey scale images need to be segmented prior to the mainstream procedure. However, segmentation of MR images is a problematic task [11] because of their properties (mentioned above) and nature of brain tissue (it is difficult to precisely define borders between extremely complex biological structures). Besides, the tissue to be analysed has to be designated (usually white matter, grey matter or both). This procedure is both time-consuming and also has a significant impact on the final result[10]. In the proposed framework, a variation of box-counting method, proposed by Sarkar and Chaudhuri (differential box-counting)[12], has been implemented. It was chosen because it operates directly on grey scale images and does not depend on a special preprocessing scheme.

An image of size $M \times M$ is scaled down to $s \times s$. Then $r = s/M$. The 2D image is treated as a 3D image, where (x, y) denotes 2D position and z denotes a grey level. A column of boxes $s \times s \times s'$ is obtained. If the total number of grey levels is G , then $G/s' = M/s$. If the minimum and maximum grey levels of the image in the grid (i, j) fall in the box k and l respectively, then [12]

$$n_r = l - k + 1 \quad (2)$$

is a contribution of the grid (i, j) to N_r :

$$N_r = \sum_{i,j} n_r(i, j). \quad (3)$$

N_r is calculated for various values of r as in simple box-counting. Fractal dimension D is then estimated from least square linear fit of $\log(N_r)$ against $\log(1/r)$.

3 Results

The images the procedure was tested with were acquired in Lublin (Poland) hospitals. Some of them were previously used for fractal feature analyses of formerly segmented images, described in [10]. The images were divided by an experienced radiologist into two groups: normal cases (30 subjects) and images with pathological changes related to dementia (17 subjects). Calculations were performed on intermediate slices with the most circular brain shape and maximal expansion. A few of them are shown in Fig. 1.

Fractal dimension was calculated with the differential box-counting method [12]. The calculations of N_r was performed for box sizes of: 2, 4, 8, ..., 256. The result of a linear fit of $\log(N_r)$ against $\log(1/r)$ for one of the images is shown in Fig. 2. The linear regression standard error is 0.1 for this example. It can be reduced to 0.04 by removing calculations for box sizes of 2 and 256.

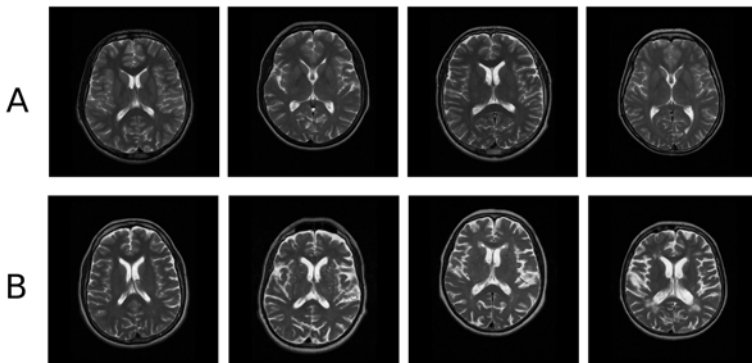


Fig. 1. Examples of the test images: A – normal cases, B – pathological cases

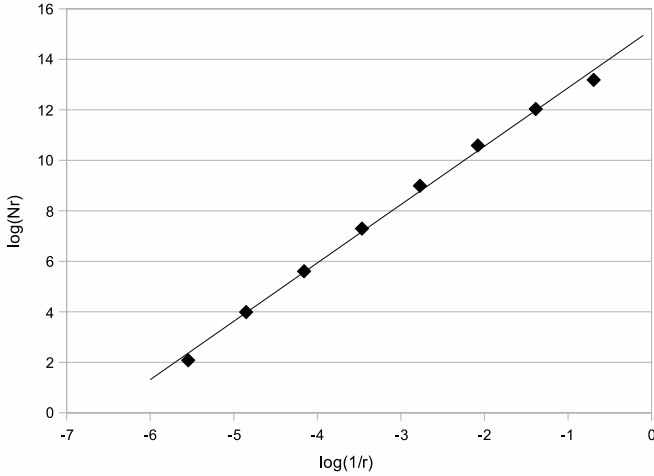


Fig. 2. $\log(N_r)$ versus $\log(1/r)$ for one of the test images

Fractal dimension calculation results are presented in Fig. 3. It is observable that fractal dimension for normal images is significantly lower than in case of pathological changes. Average value for normal images $\overline{D_{norm}} = 2.27$, for pathological images: $\overline{D_{abnorm}} = 2.32$. Normal values lie between 2.24 and 2.30. Abnormal ones are located between 2.30 and 2.35.

No strict criterion was used for election of the slice to be analysed (an intermediate slice with the most circular brain shape and maximal brain expansion), so it is desirable to analyse the influence of the election process on the fractal

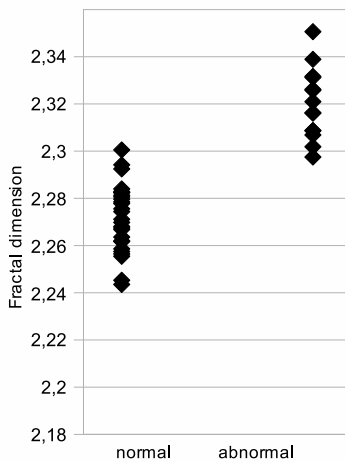


Fig. 3. Fractal dimension of normal and abnormal MR brain images

dimension calculation. The results for two examples are presented in Fig. 4. The region of interest was located between slices 15 and 20. It is observable that fractal dimension is stable in this slice ranges for all tested image datasets.

The fractal dimension calculation is relatively fast. For a single 512×512 slice it takes about 0.5s on a typical PC (Intel Core2 Duo, 1.5GHz, Linux kernel 2.6.22, program written in C++, gcc 4.1.3 compiler).

4 Discussion

According to calculation of linear regression standard error and results presented in Fig. 2, fractal dimension was found reliably for MR brain images, with the differential box-counting method, though it was not the main purpose of this work. It is noticeable that a very clear borderline between normal and abnormal cases is visible in Fig. 3 presenting results of fractal dimension calculation. Only a general classification (normal and abnormal cases with different kinds of neurological disorders), regardless of age, gender, etc. was performed. However, it has been proved that fractal dimension is a robust measure that can be applied for delineation between radiological images of normal and pathological brain structures.

The presented procedure operates directly on unprocessed MR images. In particular, no image segmentation (a baffling task) and no brain extraction was performed. Despite that the final result is satisfactory (almost faultless test image classification).

Selection of a slice (or slices) to be processed seems to be a problematic issue. Though, it was observed that in the brain region of interest (slices 15–20 in Fig. 4) fractal dimension value is very stable across a few neighbouring slices. Accurate selection of the right slice is not critical. Fractal dimension can also be

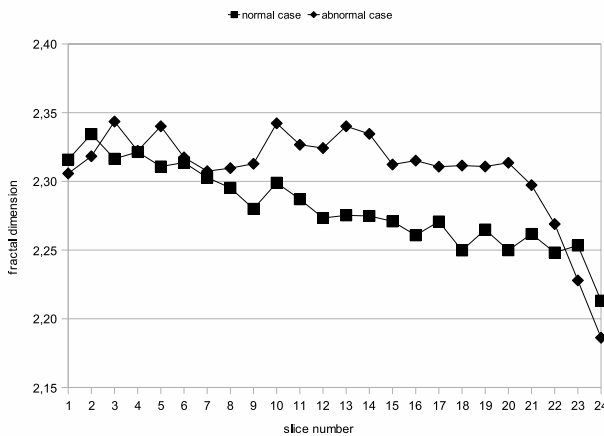


Fig. 4. Fractal dimensions for all individual slices of a single dataset (for one normal and one pathological dataset)

calculated for many slices (which is not a problem because it requires only 0.5s per slice on a moderate PC) and the interesting range of slices can be located automatically. Also presence of non-brain tissue in an image does not cause a significant impact on the classification result.

The next step is to classify also abnormal images regarding different kinds of pathologies. The authors are also going to introduce external knowledge (age, gender, demographic information etc.) into the classification framework, by means of fuzzy logic. An attempt of fuzzy fractal-dimension-based segmentation of pathological structures is also going to be made.

References

1. World Health Organization, International Statistical Classification of Diseases and Related Health Problems. 10th Revision (ICD-10), Mental and behavioural disorders, ch.V, F00–F07 (2007)
2. Castrejón-Pita, A.A., Sarmiento-Galán, A., Castrejón-Pita, J.R., Castrejón-García, R.: Fractal Dimension in Butterflies Wings: a novel approach to understanding wing patterns? *Journal of Mathematical Biology* (2004) DOI: 10.1007/s00285-004-0302-6
3. Yua, L., Zhangb, D., Wang, K., Yanga, W.: Coarse iris classification using box-counting to estimate fractal dimensions. *Pattern Recognition* 38, 1791–1798 (2005)
4. Xu, S., Weng, Y.: A new approach to estimate fractal dimensions of corrosion images. *Pattern Recognition Letters* 27, 1942–1947 (2006)
5. Kiselev, V.G., Hahn, K.R., Auer, D.P.: Is the Brain Cortex a Fractal? *Sonderforschungsbereich* 386, Paper 297 (2002)
6. Majumdar, S., Prasad, R.: The fractal dimension of cerebral surfaces using magnetic resonance imaging. *Comput. Phys.*, 69–73 (1988)
7. Mandelbrot, B.B.: *Fractal Geometry of Nature*. Freeman, San Francisco (1982)
8. Pentland, A.P.: Fractal based description of natural scenes. *Trans. Pattern Anal. Machine Intell, PAMI* 6, 661–674 (1984)
9. Bisoi, A.K., Mishra, J.: On calculation of fractal dimension of images. *Pattern Recognition Letters* 22, 631–637 (2001)
10. Buczko, O., Mikołajczak, P.: Fractal feature analysis of the human brain structures in neuroanatomy changes. *Annales UMCS Informatica AI* 5, 161–169 (2006)
11. Pham, D.L., Xu, C., Prince, J.L.: Current methods in medical image segmentation. *Annual Review of Biomedical Engineering, Annual Reviews* 2, 315–337 (2000)
12. Sarkar, N., Chaudhuri, B.B.: An efficient differential box-counting approach to compute fractal dimension of image. *IEEE Trans. Systems, Man, Cybernet.* 24(1), 115–120 (1994)

Application of MLBP Neural Network for Exercise ECG Test Records Analysis in Coronary Artery Diagnosis

Kamil Stefko

Institute of Precision and Biomedical Engineering, Faculty of Mechatronics,
Warsaw University of Technology, Sw. A. Boboli 8, 02-525 Warsaw, Poland
kamil@mchtr.pw.edu.pl

Summary. Atheromatous narrowing and subsequent occlusion of the coronary vessel cause coronary artery disease. Application of optimised feed forward multi-layer back propagation neural network (MLBP) for detection of narrowing in coronary artery vessels is presented in this paper. The research was performed using 580 data records from traditional ECG exercise test confirmed by coronary arteriography results. Each record of training database included description of the state of a patient providing input data for the neural network. Level and slope of ST segment of a 12 lead ECG signal recorded at rest and after effort (48 floating point values) was the main component of input data for neural network. Coronary arteriography results (verified the existence or absence of more than 50% stenosis of the particular coronary vessels) were used as a correct neural network training output pattern. More than 96% of cases were correctly recognised by thoroughly verified MLBP neural network. Leave one out method was used for neural network verification so 580 data records could be used for training as well as for verification of neural network.

1 Introduction

Coronary artery disease manifests as angina, silent ischaemia, unstable angina, myocardial infarction, arrhythmias, heart failure and sudden death. It causes severe disability and more death than any other disease including cancer. Coronary artery disease is due to atheromatous narrowing and subsequent occlusion of the coronary vessel. An abnormal electrocardiogram increases the suspicion of significant coronary disease, but a normal result does not exclude it. Therefore the exercise electrocardiography is the most widely used non-invasive test in evaluating patients with suspected angina. It is generally safe method and provides diagnostic as well as prognostic information. The average sensitivity and specificity is 68% and respectively 77% [5]. The test is interrupted in terms of achieved workload, symptoms and electrocardiographic response. A 1 mm depression in the horizontal ST segment is usual cut-off point for significant ischaemia. Poor exercise capacity, an abnormal blood pressure response and profound ischaemic electrocardiographic changes are associated with a poor prognosis [1].

In this paper application of feed forward Multi-Layer neural network (NN) trained with Back Propagation algorithm (MLBP) applied to interpretation of

results of traditional ECG exercise test is presented. The database used for training and verification of neural network contains 580 data records from exercise test (training input data) and results of coronarography as a pattern of correct answer of neural network (training output data).

Application of leave one out verification method is a very important aspect of this experiment. This method gives the best generalisation estimate but due to very high computation demand is applied for small datasets only.

2 Materials and Method

2.1 Neural Network

Neural network methods attempt to define decision regions in feature space, according to their interconnection of weights within the network. The weights are determined in the iterative training phase, during which samples are presented to the NN input. After passing through the network, resulting signal is compared with the desired output to obtain an error expression, which is then back-propagated and used as a factor to correct weights. Error is iteratively minimised until the network converges on the solution.

In this experiment, feed forward perceptron type neural network was implemented (Fig. 1). Artificial neural networks consist of simple processing units, called neurons (S), and weighted connections between them (w).

In a feed forward multilayer perceptron architecture the neurons are arranged in layers and a neuron from one layer is fully connected only to each neuron of the next layer. The first ($S(0)$) and last ($S(L)$) layer are the input respectively output layer. The layers between them are called hidden. Values are given to the neurons in the input layer and the results are taken from the output layer. The outputs of the input neurons are propagated through the hidden layer of the net. Figure 2 shows the schematics of algorithm each neuron performs.

Neural network trained with back propagation algorithm with momentum term and adaptive learning rate was especially optimised. Neural network has one hidden layer, 60 input units, and four decision neurons in the output layer to indicate the state of particular coronary vessel.

The optimum configuration of feedforward multilayer perceptron network with its input, hidden and output layers are very difficult to find. Too many hidden neurons cause disability of neural network to extract the function rule and take more time for learning. With a lack in hidden neurons it is not possible to reach any error limit. Input and output layers are determined by the problem and the hidden layer neuron count and determined by the function that is to be approximated. Finally neural network with 5 neurones in the hidden layer (60-5-4 architecture) was applied.

The basic back propagation algorithm was used with added momentum term which allows the network to ignore small changes in the error surface. Beside that, the adaptive learning rate was used which attempts to keep the learning step as large as possible while keeping learning process to converge to solution.

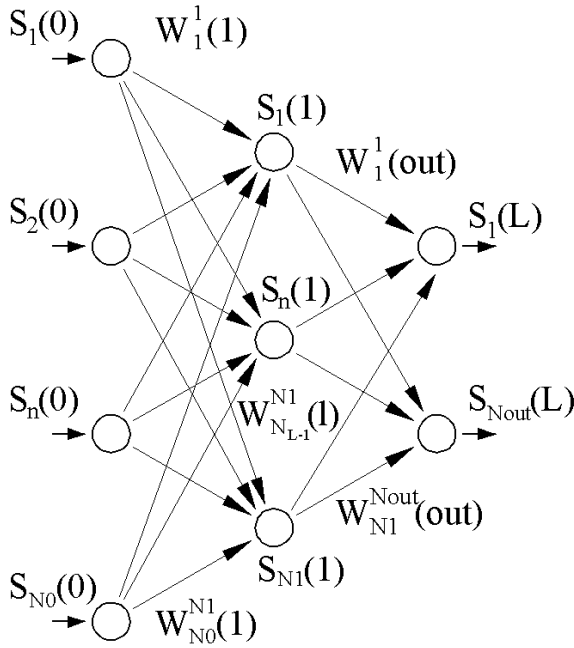


Fig. 1. Schematic view of the architecture of the neural network

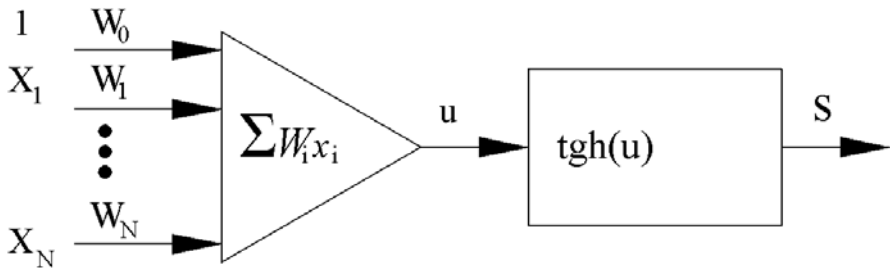


Fig. 2. Schematic view of the architecture of the single neuron

2.2 Medical Database

The research was performed using 580 data records from traditional ECG exercise test confirmed by coronary arteriography results. Each record of training database contained description of the state of a patient, provided input data for the neural network. Level and slope of ST segment of a 12 lead ECG signal recorded at rest and after effort (48 floating point values) was the main component of input data for neural network.

Coronary arteriography results of verified existence or absence of more than 50% stenosis of the particular coronary vessels were used as a correct neural network training output pattern. Due to non-specific results of ECG exercise test result of female patient, only male patient data were included in training database.

Detailed description of the applied medical database is described by Lewenstein in [4]. In this work generalisation possibility of different NN architectures was evaluated. Neural network with MLBP architecture was the first one investigated. This network obtained correctness of simple diagnosis equal 88% and 75% correctness of patient state diagnosis.

3 Results and Discussion

One of the most important aspects of NN application is how well the NN generalises to unseen data. The best generalisation possibility has that one neural network that offers the smallest probability of misclassification, which can be estimated by the ‘leave one out’ (LOO) method. This method is based on classification of each input record from the database by neural network trained with the whole database reduced by classified record. Thus the number of NN training sessions equals to the numerical force of the reference set.

Generalisation error is better for smaller nets. They need more epochs to learn the rule of coronary artery disease, but because of this they can generalise their behaviour better. Beside that, with neural network with lower neurone count in the network, it was very difficult to obtain trained network. Training algorithm will not converge to result within desired number of training epochs.

Three hundred neural network training experiments were performed with fixed, optimum neural network architecture (60-5-4) and fixed training parameters. While majority of training experiments failed to finish training process, for clinical application only one neural network is needed, that one which was marked by the best generalisation possibility. In Table 1, generalisation possibility as well as standard deviation computed for five best performing neural networks is shown.

As a main indicator of neural network quality, the PST (Patient State) parameter was used. This parameter is a generalisation possibility of the one neural network that classifies four coronary arteries as narrowed or healthy.

Table 1. Average generalisation possibilities of applied neural network estimated with live one out method

Diagnosis type	SD	PST	SENS	SP	NARROWED			
					LCA	LAD	LCX	RCA
Generalisation possibility (%)	99.17	96.58	100.0	75.83	99.17	99.51	98.65	99.20
Standard deviation	1.30	1.83	0.00	37.74	1.30	0.33	1.14	0.84

Tested with database of 580 medical cases, the 96% of probability of proper indication was obtained. This result proved the thesis that in ECG exercise test data there is information about the narrowing in the particular coronary artery vessels. Cardiologists said that there is no enough information in results of ECG exercise test for such a diagnosis, and they do not make it.

All of the other indicators shown in Table 1 were computed for reference only. Those parameters are sensibility (SENS), specificity (SP) and simple diagnosis (SD). It is known from medical literature that sensitivity of the exercise test is from 44 to 88%, average 68% and specificity is from 52% to 89%, average 77% [2, 5]. The result of this research of 100% sensibility indicates increase of 32%. The lower specificity obtained in this experiment (76%) is due to insufficient count of healthy patient data in training base.

SD parameter is just a simple classification of a patient as sick or healthy. Patient is classified as a sick while at last one coronary vessel is narrowed. Probability of correct diagnosis of 99% is a very good result. The last four entries in the table represent possibility of four main coronary arteries (Left coronary artery (LCA), Left anterior descending artery (LAD), Left circumflex artery (LCX) and Right coronary artery (RCA)) to be classified as a narrowed or healthy independent each of the other.

4 Conclusion

Our experiment confirmed that it is possible to localize narrowing in the particular coronary artery of four main coronary vessels with aid of MLBP neural network applied for traditional ECG exercise test interpretation. The obtained result of 96% of probability of the correct answer of MLBP neural network introduces 21% improvement to the result obtained by [4] for the same neural network architecture.

Acknowledgement. Computations were performed at the Interdisciplinary Centre for Mathematical and Computational Modelling (ICM) of Warsaw University using Cray SV1ex parallel vector supercomputer (Grant number G12-6).

References

1. Froelicher, V.: Exercise Tests Manual. Bell Corp, Warsaw (1999)
2. Gianrossi, R., et al.: Exercise included ST depression in the diagnosis of coronary artery disease: a metaanalysis. *Circulation* 80, 87–98 (1989)
3. Grech, E.D.: ABC of interventional cardiology: Pathophysiology and investigation of coronary artery disease. *BMJ* 326, 1027–1030 (2003)
4. Lewenstein, K.: Artificial neural networks in the diagnosis of coronary artery disease based on ECG exercise tests. *Ofcyna wydawnicza PW, Electronics* 140, 53–57 Warsaw (in Polish) (2002)
5. Opolski, G.: Choroba niedokrwienna serca. In: W. Januszewicz, F. Kokot: *Interna*, PZWL 135-177 Warsaw (in Polish) (2001)

Volumetric Analysis of Tumours and Their Blood Vessels

Rafal Henryk Kartaszynski¹ and Pawel Mikolajczak²

¹ Maria Curie Sklodowska University, Marii Curie-Sklodowskiej 1 square, 20 - 031 Lublin, Poland

hatamoto@goblin.umcs.lublin.pl

² Maria Curie Sklodowska University, Marii Curie-Sklodowskiej 1 square, 20 - 031 Lublin, Poland

mikfiz@goblin.umcs.lublin.pl

Summary. In this paper we present first results of our, computer aided, research into angiogenesis process, conducted in association with radiologists from local clinical hospital. Presented here is its informatics part, which was to estimate, basing on CT scans, the quotient of the tumour volume to the number of its capillary veins. Should some correlation be found between the effectiveness of the cancer healing process and the quotient, it would mean that healing is affecting the angiogenesis process.

Angiogenesis is a process of forming new blood vessels from the already existing ones, and is the main cause of violent cancer development. Let us say that cancer cells force the angiogenesis process, thus making vasculature nourishing cancer cells a colony and enabling its growth. Without it, a tumour would be harmless. It is no wonder that modern medicine is trying hard to develop a method to stop the angiogenesis process. If during treatment the number of capillaries decreases, the tumour is less effectively nourished and the disease recedes, furthermore the ability of the cancer to spread over the body is being limited.

1 Introduction

The switch in tumours from the quiescent state to malignancy is signalled by the commencement of the angiogenesis process. tumours need an extensive network of capillaries to provide nutrients and oxygen. solid tumours will not grow beyond 2 millimetres without new blood vessels. malignancy and invasion are angiogenesis dependent. that is why, the obvious idea would be to stop angiogenesis, and thus block cancer growth. therefore, this process is the subject of extensive research and may lead to the discovery of a remedy for some tumours.

Question 1. *Is there any relationship between the number of capillary veins supplying tumour and the progress of cancer treatment?*

Radiologists, from the Department of Radiology at SPSK Hospital No 1 in Lublin (Poland), who put forward this idea, were not sure of the answer to the above question and could not enumerate any academic paper where it might have been analyzed, and suggested that this issue could be a subject of clinical research.

Analysis of numerous patients' treatment processes could answer the question: does effective treatment cause the number of tumour veins to decrease, thus limiting the cancer's ability to grow and spread?

2 Methodology

Computed Tomography pictures before and after application of contrast medium will be analyzed. As we know CT is based on X-ray, therefore after injection of contrast medium into the patient's body, blood vessels carrying it will be highlighted on a CT scan. This enables us to choose (of course by adequate processing of CT images) only vessels belonging to a specific organ (or tumour). At this point another question arises: is it possible to find such a small object as tumour vasculature on CT scans (is CT spatial resolution large enough to distinguish them from other tissues). The answer is ambiguous: they can be distinguished, but almost certainly not all of them. Fortunately rapid development of modern CT will soon provide physicians with new, more accurate and effective diagnostic devices with larger spatial resolution, allowing detection of the smallest capillaries. Therefore, we can be sure that technical means for realising research topic are adequate and the only thing left to do is to put research into effect.

As mentioned above we are cooperating with the local clinic, and our contribution to the research effort was to implement a computer program which analyzes medical images, and allows physicians to measure and examine the proportions between tumour volume and the number of its capillary veins. The clinical research part using this program has the following structure:

- Taking CT scans of patients' body parts of interest to us before application of contrast medium.
- Taking CT scans of patients' body parts of interest to us after application of contrast medium. Amount of contrast medium will be carefully selected to ensure its smooth propagation along blood circulation system and objectivity of examination results.
- Segmentation of a specific tissue - tumour from scans before or after contrast medium application. Data from which cancer will be segmented depends mainly on the kind of tissue. If it is easier to segment it after application of contrast medium, a second data set will be used, and vice versa.
- Estimation of tumour volume according to the data segmented before.
- Segmentation of tumour capillary veins.
- Estimation of the number (and volume) of capillary veins supplying the tumour
- Determination of the quotient of tumour volume to the number of its capillary veins

2.1 Tumour Segmentation

The presented segmentation method is developed mainly to solve the issue of segmentation of various tumour types. That is why we may make some preliminary

assumptions, regarding, for example, tissue three-dimensional shape, its features (for example small density changes in tumour volume). However, these assumptions are not needed for the algorithm in its basic shape. This information may be necessary while modifying the basic algorithm for three dimensional cases. Another thing is that we do not require the algorithm to be very quick, we would rather prefer it to be accurate (the presented approach can ensure this point - as experiments have shown).

The method is based on cellular automata, firstly introduced by Ulam and von Neumann in 1966 [1] and is an adaptation of idea presented in [2]. It can be used to solve difficult segmentation problems, furthermore it is multilabel - segments many object simultaneously (computation time does not depend on the number of labels). It is interactive: requires the user to provide starting points for the algorithm (not many seeds are needed and their entering is not laborious), but in turn enables him to observe the segmentation process and make modifications in it. Interactivity is very important for physicians who like to have some (often large) influence on medical images processing. Furthermore, a radiologist will be able to place seed points very accurately and in characteristic places of a specific organ, and will check the correctness of segmentation afterwards.

The method is extensible, allowing simple modification of the algorithm for a specific task. As it was shown in our previous work [3, 4] it is very accurate for two- and three- dimensional medical images. Three-dimensional cases require some unsophisticated data post processing, or making some modifications in the manner in which the automaton grows into the third dimension from the two-dimensional layer.

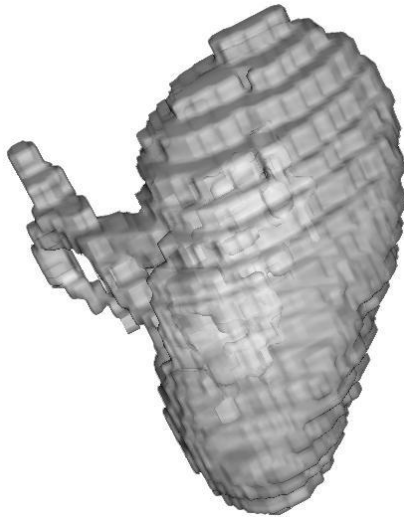


Fig. 1. Kidney segmented using described method

2.2 Tumour Volume Estimation

After segmentation of a specific organ (tissue tumour) we obtain a mask of this object. This mask can be then used to extract a given part of the three dimensional image of the patients' body after application of contrast medium. This way we now operate only on an image containing only the tumour. Knowing the number of voxels belonging to the object we can easily estimate its volume, by simply multiplying it by the volume of a single voxel. This last value is characteristic of a specific image and is written in its DICOM file.

Of course, the obtained value is only an estimation, because, as it is easy to foresee, no segmentation method is perfect, and the segmented object can vary from the original (ideal one) by a few percent. This is natural but should fit in statistical error. Another fact showing that this can be only an estimation is that, CT spatial resolution is limited. That is the border of the organ only in few cases will go exactly between two voxels, one of which belongs to the object and the other to its surrounding. In most cases the boundary goes through the voxel but CT interprets it as an object voxel (if the object is in majority in this box - voxel) or surrounding voxel (if the object is in minority). But all these details do not have a great influence on the final result and can be omitted.

2.3 Capillary Veins Segmentation

Another point realized by the described program is tumour capillary veins segmentation. At this point we must remember that we deal with images after application of contrast medium, and that veins should be highlighted. Contrast is injected into a vein, flows through it into the heart and to the aorta. Unfortunately, after going through the arteries and capillaries it reaches the organs' parenchyma and is absorbed there. As we see contrast medium flowing through veins must be distinguished from the one absorbed by parenchyma. Density (intensity) value corresponding to blood mixed with contrast medium in tumour capillaries should be the same as the one flowing through the nearest artery and even the aorta. Therefore, detection of capillaries is realized by finding voxels of specific density in the tumour volume. In most cases it is not enough and some post processing must be done. This post processing consists of application of the opening filter (size of the kernel and number of operations must be chosen for the specific individual case), which removes all information noise.

The person using this program must define the density of reference for the above process. This can be done in two ways. Firstly, by clicking the selected pixel on the displayed image and providing a tolerance factor (in per cent) which states the bracket of accepted density values. Tests have shown that about 3 per cent is a good choice. Secondly, by circling the part of the image (containing for example a section of the aorta with contrast medium) from which the average value is chosen. Distribution of density in this circle should resemble in shape the Gaussian function with a lot of values near the average value and a small number of intensities varying significantly from it. These diverging values may be a kind of noise and should be omitted. Therefore, it is possible to set brackets

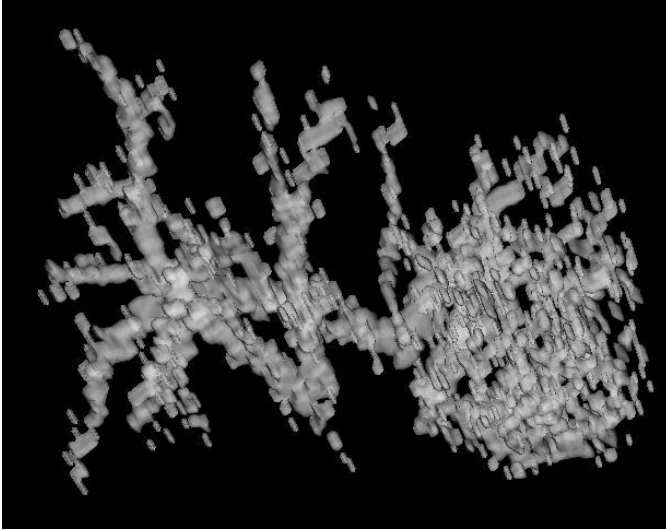


Fig. 2. Visualisation of segmented capillaries in liver

(in per cent of the range of values from the average to the outermost value) for accepted density. Below we present results of such approach - segmented liver blood vessels.

2.4 Capillary Veins Counting

After segmenting veins we must estimate their number and volume (this information may be useful for a radiologist). While the second point is easy and proceeds similarly to counting the volume of the tumour, the first is not that trivial. Its realization requires use of the connected components labelling algorithm.

Before analyzing this feature of the research, let us define how we understand the term: "one capillary vein". As we know capillaries are formed in a shape resembling a tree. It is important to state whether one capillary is a whole tree or maybe only its branch (if so: where it starts, where it ends?). We have assumed that one capillary is a whole tree with its root on the boundary of a tumour.

Now it is easy to understand how the counting process is realized. Three dimensional connected components labeling is launched [5]. As a result we will receive groups of voxels divided into several differently labelled groups. Some of these groups may contain some kind of 'noise', that is voxels identified as vessels but not connected with capillaries. Therefore it is essential to exclude them. In order to do this all labelled groups without common voxels with the object boundary must be removed from the deliberations. To do so, the algorithm is first tracking the three dimensional boundary of the analyzed object [6, 7]. Secondly, for all labelled groups it checks if the current group has a voxel common with the boundary, if it has not, then it is removed. Now the only thing left to do is to count the remaining groups. Of course, in the implementation it is

done differently. When the connected components analysis is done, we know the number of the assigned labels, and when a labelled group is removed, this number decreases, giving in result the number of the capillary veins of the tumour.

2.5 The Quotient of the Tumour Volume to the Number of Its Capillary Veins

When one knows the volume of the tumour (TV) and the number of its blood vessels (NBV), the estimation of their quotient (Q) is trivial, the first value must be divided by the second.

$$Q = \frac{TV}{NBV} \tag{1}$$

3 Results of Pilot Investigation

Unfortunately, due to the fact that clinical research is just starting, at the moment of writing this paper, we are only able to provide reader with the preliminary investigation results. In below table we present five different medical cases, provided to us by colleagues from the Department of Radiology at SPSK Hospital No 1 in Lublin. Our application was used at the beginning of the treatment to estimate the quotient of tumour volume to the number of its capillary veins (Q). As treatment continued the Q value was again calculated and compared with first result. In all cases physicians have stated that the treatment is proceeding as planed and is giving good results. In three cases tumour volume as well as Q value has decreased. In rest, volume stayed relatively unchanged while Q increased, what suggests that number of tumour blood vessels decreased.

Below you will also find, as an example, selected CT slices of the lung tumour (position 1 in *Table 1*).

General conclusion can be put forward. Indeed the answer to the question, which we stated at the beginning of this article (*Question 1*) is Yes: there is a dependency between cancer treatment and quotient Q . However more detailed and thorough clinical research will follow, to confirm these first results and conclusions.

Table 1. Results of pilot investigation. Quotients Q and Tumour Volumes were calculated with our application

no.	tumour name	Q before ($TV [cm^3]$)	Q after ($TV [cm^3]$)
1	small-cell lung cancer	0.55 (3.85)	0.71 (3.25)
2	astrocystema fibrillare	0.68 (4.12)	0.79 (3.64)
3	careinama adenoides custicum	0.62 (4.88)	0.77 (3.12)
4	careinama pancreatis	0.56 (5.10)	0.83 (5.00)
5	careinama adenoides cysticum	0.7 (4.23)	1.02 (4.10)

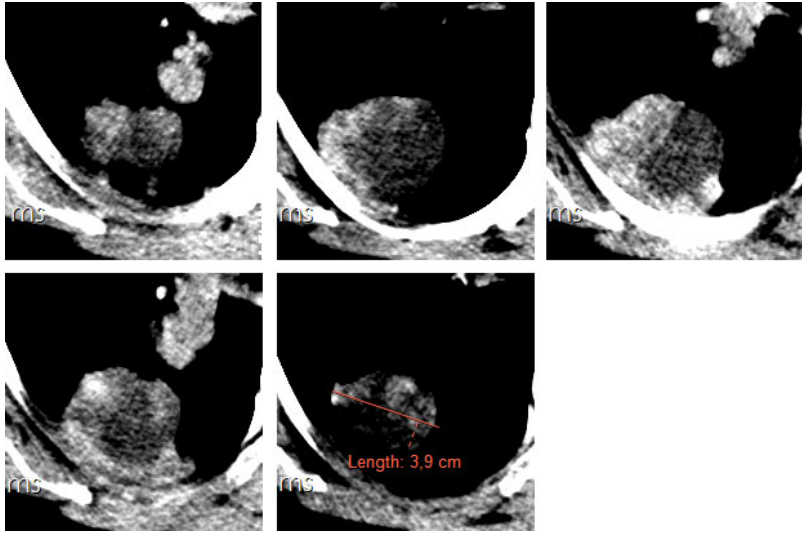


Fig. 3. Example of lung tumour CT slices

4 Conclusions and Future Work

The aim of the whole project is to research into tumours angiogenesis by analyzing the quotient of the tumour volume to the number of its blood vessels. This quotient must be estimated by a computer program analyzing CT images of the patient's body part affected by the cancer. This issue of the whole research, was presented in this article. We have discussed general methodology of acquisition, processing and analysis, of these images. Whole methodology has been implemented as Windows application, thoroughly tested, and is now being used in clinical research. First results of this research has been shown and described in section 3 of this article.

First results are very promising. According to them we can state that there is a correlation between progress of the treatment and the quotient of the tumour volume to the number of its capillary veins. These are preliminary results and should be confirmed during further clinical research.

References

1. Von Neumann, J.: Theory of Self-Reproducing Automata. University Of Illinois Press. Ed. And Completed by A. Burks (1966)
2. Popovici, A., And Popovici, D.: Cellular automata in image processing. In: Fifteenth International Symposium on Mathematical Theory of Networks and Systems (2002)
3. Kartaszynski, R., Mikołajczak, P.: CATS - Cell Automata based Tissue Segmentation of two- and three-dimensional CT and MRI medical images. Annales Informatica, UMCS, Lublin (2006)

4. Kartaszyński, R., Mikołajczak, P.: CABRS - cellular automaton based MRI brain segmentation. In: Proceedings o the XI International Conference MIT 2006, Wisła - Malinka (2006)
5. Park, J.M., Chen, H.C.: Fast Connected Component Labeling Algorithm using a divide and conquer technique. University of Alabama, Tuscaloosa (2000)
6. Marr, D., Hildreth, E.C.: Theory of edge detection. Proc. R. Soc. London Ser. B (1980)
7. Lee, J.S.L., Haralick, R.M., Shapiro, L.S.: Morphologic Edge Detection. In: 8th International Conference on Pattern Recognition. IEEE Computer Society, Paris (1986)

Pre- and Postprocessing Stages in Fuzzy Connectedness-Based Lung Nodule CAD

Paweł Badura and Ewa Piętka

Silesian University of Technology, Institute of Electronics
pawel.badura@polsl.pl, ewa.pietka@polsl.pl

Summary. The crucial part of the lung cancer computer-aided diagnosis (CAD) is the segmentation of pulmonary nodules in Computed Tomography (CT) study. A new multilevel approach based on fuzzy connectedness principles has been developed. The three-dimensional fuzzy connectedness analysis requires a dedicated preprocessing stage in order to limit the computation time to a reasonable range. It consists of the initial thresholding, connected components labeling, and creating the binary masks of regions within the thorax. For nodules connected to pleura or vessels, a separation step is needed, using mathematical morphology and the shape analysis. Separation of the nodule and pleura is performed in the preprocessing stage, whereas separation of a nodule and connected vessels – in the postprocessing stage. In this paper the methodology is described and illustrated.

The whole segmentation method has been tested on a set of three-dimensional CT images of the thorax with delineated lung nodules. Results and some examples of such an application are shown.

1 Introduction

Lung cancer has a very high mortality rate, causing ca. 1.3 million deaths per year, and being one of the most dangerous diseases [11]. The main reason of its progress is cigarette smoking. The curability rule is clear: the earlier and more accurate the diagnosis, the higher the survival chance for the patient. So, the biomedical imaging-based computer-aided diagnosis is of great importance.

Three imaging techniques are used for lung cancer diagnosis. The thorax RTG has been used in the 80's of the twentieth century, whereas the Positron Emission Tomography (PET) is still a matter of future, mainly due to economical conditions. Since 20 years Computed Tomography has been the main biomedical imaging technique. Today's CT scanners give a submillimeter slice thickness and pixel spacing, and allow for a precise three-dimensional analysis of anatomical structures.

Many segmentation or detection methods have been developed for pulmonary nodules [4]. In this paper, a multilevel approach based on fuzzy connectedness [10, 9] is described. It also uses an artificial intelligence, mathematical morphology, and the shape analysis. Some of its components have already been presented in earlier works [2, 3]. The blocks being circumstantially discussed here, are the pre- and postprocessing stages.

This paper is organized in the following manner. First, the lung nodule segmentation scheme is presented in section 2, along with a brief description of those blocks, which have been presented in earlier works. In sections 3 and 4 the pre-processing and postprocessing stages are shown, respectively. Finally, section 5 includes the results obtained during the evaluation of the whole segmentation method along with conclusions.

2 Lung Nodule Segmentation Scheme

Fig. 1 shows the workflow of the presented lung nodule segmentation method. It consists of four stages: (1) interactive seed points selection, (2) preprocessing, including next two blocks, (3) main segmentation stage, with an automatic selection of the object and background seed points and three-dimensional fuzzy connectedness analysis, and (4) postprocessing including the last block. The shaded blocks are investigated further in this paper, the remaining are briefly described here.

First, two seed points are marked manually by the radiologist: \mathbf{o} representing the object (nodule) and \mathbf{b} – the background (lung parenchyma). Further analysis is performed automatically. The preprocessing stage, described in section 3, is followed by two procedures of automatic selection of a larger number of seed points within both, object and background. These algorithms have been presented in an earlier work [3]. Then, the three-dimensional fuzzy connectedness analysis is implemented. The fuzzy connectivity scene [10, 9] is computed using Dijkstra’s algorithm [2]. The algorithm terminates, when the first background seed point is reached, resulting in thresholding the fuzzy connectivity scene [2]. Finally, the binary object is processed, in order to eliminate connected vessels and make some correction operations.

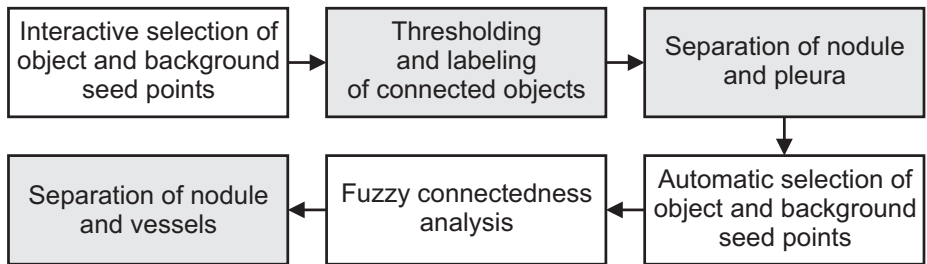


Fig. 1. Scheme of the lung nodule segmentation process

3 Preprocessing Stage

3.1 Thresholding and Labeling of Connected Objects

The main goal of the preprocessing stage is a separation of the nodule and pleura. These regions are marked by two binary masks M_n and M_a , representing the

nodule and regions surrounding the lungs, respectively, and defined as:

$$\forall \mathbf{c} \in C^3 : M_i(\mathbf{c}) = \begin{cases} 1 \Leftrightarrow \mathbf{c} \text{ belongs to region } i \\ 0 \Leftrightarrow \text{otherwise,} \end{cases} \quad (1)$$

where C^3 is a set of image voxels (the image domain) and i stands for n or a .

A CT thorax image I is first binarized, using Otsu thresholding method [6]. A threshold value t_{Otsu} , computed for the entire image, yields a binary image I_{Otsu} :

$$\forall \mathbf{c} \in C^3 : I_{Otsu}(\mathbf{c}) = \begin{cases} 1 \Leftrightarrow I(\mathbf{c}) > t_{Otsu} \\ 0 \Leftrightarrow I(\mathbf{c}) \leq t_{Otsu}, \end{cases} \quad (2)$$

Thus, I_{Otsu} contains two classes: the object with white voxels, and the background with black voxels. The object includes mediastinum, diaphragm, bones, skin, pleura, vasculature etc., whereas the lung parenchyma, airways, air around the patient are assigned to the background. In most cases the pulmonary nodule also belongs to the object¹.

All white voxels in I_{Otsu} form a set of three-dimensional, distinct connected components, depending on the adjacency relation. Here, based on 26-adjacency relation ϱ_{26} :

$$\forall \mathbf{c} = (c_x, c_y, c_z), \mathbf{d} = (d_x, d_y, d_z) \in C^3 \quad \mathbf{c} \varrho_{26} \mathbf{d} \Leftrightarrow \max_{i=x,y,z} \{|c_i - d_i|\} \leq 1, \quad (3)$$

connected components are detected and labeled with consecutive integers.

A connected component l_a with the largest number of voxels (volume) is the one including all regions surrounding the lungs, so it also includes the pleura. A connected component including seed point \mathbf{o} (and a nodule itself) is labeled with l_n . Now it is the right time to introduce the types of pulmonary nodules considered in the current study [5].

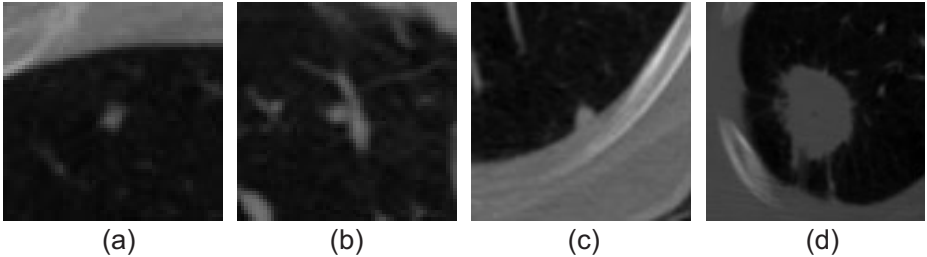


Fig. 2. Pulmonary nodule types: (a) well circumscribed, (b) vascularized, (c) juxta-pleural, (d) "pleural tail"

¹ If it does not ($I_{Otsu}(\mathbf{o}) = 0$), then no more preprocessing operations are performed. M_n consists of zeros only, and M_a corresponds to the connected component l_a within I_{Otsu} , with the largest number of voxels.

If $l_n \neq l_a$, then there is no connection between the nodule and pleura. It means, that the nodule is well-circumscribed (type A, fig. 2a) or vascularized (type B, fig. 2b). In both cases the separation block is unnecessary. Voxels labeled with l_n become 1-valued in the mask M_n , yet voxels labeled with l_a are 1-valued in the mask M_a .

If $l_n = l_a$, then the nodule is connected to pleura. The nodule can be juxtapleural (type C, fig. 2c) or the connection can have a form of a thin "pleural tail" (type D, fig. 2d). Determination of binary masks requires a separation of the nodule and pleura.

3.2 Nodule Detection

2D Connected Components Analysis

The regions are separated slice-by-slice in three steps, starting from the slice including a seed point \mathbf{o} , in both directions. First, all connected components on the slice of I_{Otsu} are labeled using an 8-adjacency relation:

$$\forall \mathbf{c} = (c_x, c_y), \mathbf{d} = (d_x, d_y) \in C^2 \quad \mathbf{c} \rho_8 \mathbf{d} \Leftrightarrow \max_{i=x,y} (|c_i - d_i|) \leq 1, \quad (4)$$

where C^2 is a set of all slice pixels (the slice domain). Components l_a^{2D} and l_n^{2D} are determined according to the rules shown above. The former indicates the region with the largest number of pixels (area), the latest – the nodule region. Similarly, if $l_n^{2D} \neq l_a^{2D}$, then there is no need to perform the separation: pixels labeled with l_n^{2D} become white in the corresponding slice of a mask M_n , yet pixels labeled with l_a^{2D} – white in the slice of M_a . This is the case in many slices with "pleural tail" nodules. However, if $l_n^{2D} = l_a^{2D}$, then the separation step is indispensable.

Breaking the Pleural Tail

The slice is processed in few steps. First, a morphological erosion operation [8] is performed on the binary slice including the l_a^{2D} component only. The structuring element is a small disk, with the diameter of a few pixels. If the connection is thin and the nodule region is large enough to survive the erosion, then there will be (at least) two connected components in the slice after the erosion (fig. 3b). If so, the nodule object is selected and morphologically dilated [8] with the structuring element used earlier by erosion (fig. 3c).

Subtraction of the nodule object from the l_a^{2D} component results in a binary slice with some small 8-connected components (few pixels each, fig. 3d), due to antiextensivity of the erosion-dilation (morphological opening) operation [7]. These components are added to the nodule region. Final partition of the l_a^{2D} component into binary masks M_a, M_n is shown in fig. 3e.

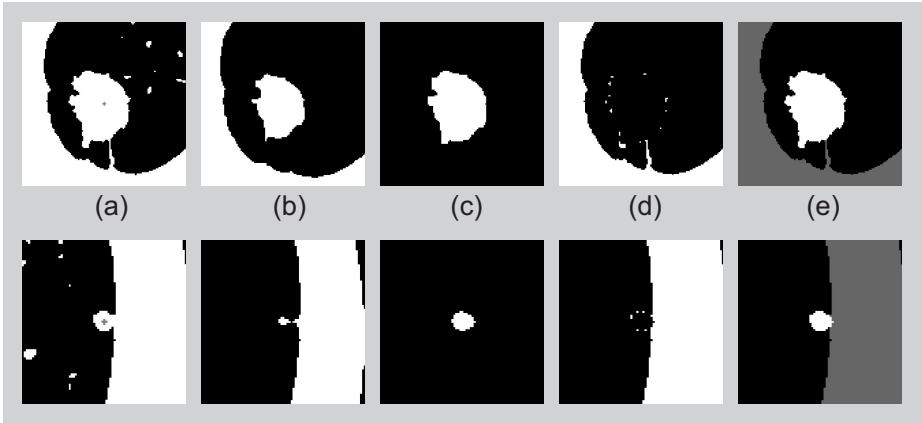


Fig. 3. Nodule detection by breaking the pleural tail (two cases); (a) ROI of a single slice of I_{Otsu} , (b) (a) after erosion, (c) nodule object selected and dilated, (d) difference between object l_a^{2D} and (c), (e) final form of M_n (white) and M_a (gray)

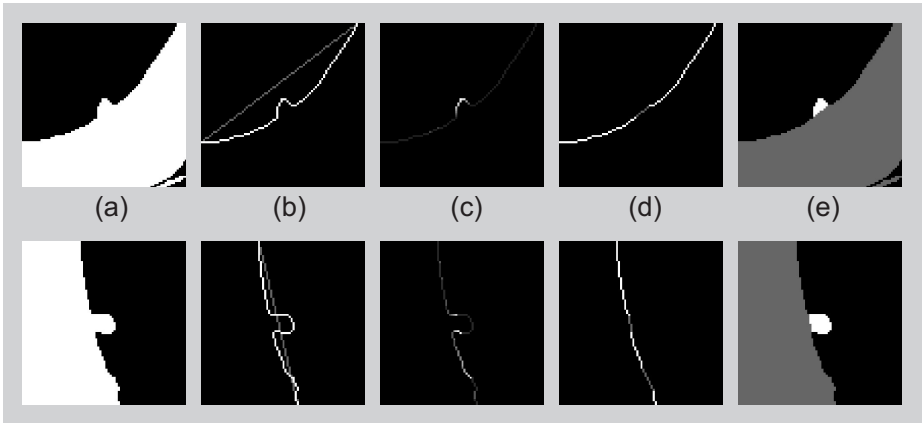


Fig. 4. Nodule detection by approximation of the lung edge (two cases); (a) ROI of a single slice of I_{Otsu} , (b) considered edge of (a) (white) and its chord (gray), (c) visualization of an auxiliary slice A , (d) approximated lung edge, (e) final form of M_n (white) and M_a (gray)

Approximation of the Lung Edge

If the nodule shares a significant amount of its surface with the pleura (juxtapleural nodule, fig. 4a), then the former step does not lead into a proper separation. Thus, an attempt to approximate the lung edge is performed.

A region of interest (ROI) of the size of $5 \times 5cm$ is chosen in a binary slice, including the l_a^{2D} component only (fig. 4a). Then, the lung edge is chosen (fig. 4b,

white line) with a local convexity bound up with the nodule. ROI-boundary edge pixels are connected by the chord (fig. 4b, gray line). Successive shift of the chord into the edge followed by the intersections results in the auxiliary slice A (fig. 4c). Based on values assigned to edge pixels in A , the edge is discontinued, and pixels of the local convexity (convexities) are removed (fig. 4d, white line). Then, elements of the partitioned edge are connected by sections (fig. 4d, gray line). Finally, the approximated edge is subtracted from the I_a^{2D} component. If that provides a successful separation of two considered regions, then corresponding slices of binary masks M_a, M_n are determined (fig. 4e). If it does not, then the entire separation stage terminates, since no more nodule region in the slice can be found.

Two binary masks M_a, M_n are used in the next two blocks of the segmentation process. Particularly, the nodule mask M_n is taken into consideration in both algorithms of automatic seed points selection, whereas all white voxels in M_a are excluded from the fuzzy connectedness analysis.

4 Postprocessing Stage

The fuzzy connectedness analysis yields a binary image I_{Co} , being a thresholded version of the fuzzy connectivity scene [2], added to a binary nodule mask M_n . It does not contain the regions surrounding the lung, because of the binary mask M_a . It might contain, however, some regions located within the lung, having intensity similar to the nodule intensity (especially vessels connected to it). A three-dimensional visualization of such a case is presented in fig. 5a. The object within I_{Co} is processed in order to remove connected vessels. Some correction operations are also performed.

In the first step, a binary image I_{fh} is obtained as a result of a morphological filling of holes in I_{Co} . Afterwards, the main part of postprocessing occurs i.e. the elimination of connected vessels employing the shape analysis and mathematical morphology. A three-dimensional distance transformation (DT) [8] of I_{fh} is computed. Then, a three dimensional morphological opening is performed on I_{fh} .

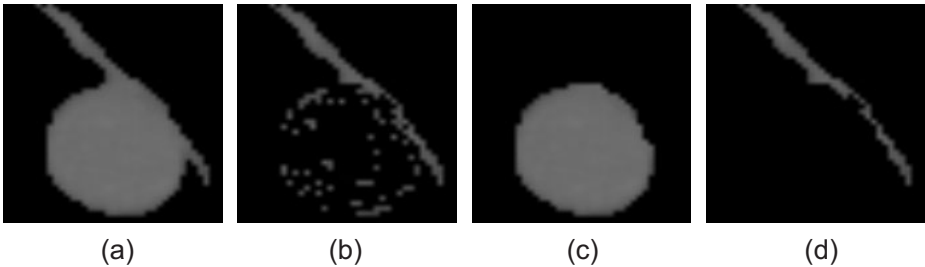


Fig. 5. 3D visualizations of consecutive postprocessing steps: (a) object after thresholding the fuzzy connectivity scene, (b) elements removed by the opening, (c) final form of a nodule, (d) elements removed after the entire postprocessing stage

The structuring element is a sphere with a diameter greater, than the maximum expected vessel diameter. It has to be noticed, that some voxels removed from I_{fh} by the opening operation belong to the nodule (fig. 5b). In order to avoid the elimination of true voxels, they are examined in following aspects: (1) their membership in 26-connected components, (2) their distance transformation values. As a result, a binary image I_{sa} is obtained. It contains no more vessels nor other undesirable regions.

Finally, the image I_{sa} is morphologically closed with a three-dimensional structuring element being a few-voxel-diameter sphere. The aim here is to smooth the edges of the object. The final form of the considered nodule and the vessel removed in the postprocessing stage are shown fig. 5c, d, respectively.

Table 1. Comparison of the *M-OB*, *Mask* and *RFC-OB* methods

	$\overline{P_{100\%}}$	$\overline{P_{50\%}}$	$\overline{N_0\%}$	$\overline{t_{pr}}$	$\overline{t_{op}}$	$\overline{t_{po}}$
Mask	93.28%	77.12%	1.18%	9.35s	0s	2.92s
RFC-OB	98.42%	83.64%	0.27%	9.35s	26.75s	1.43s
M-OB	99.94%	90.94%	0.39%	9.35s	26.75s	3.17s

5 Results and Conclusions

A set of 23 CT lung studies have been tested to evaluate the presented segmentation method. Images have been provided by the Lung Imaging Database Consortium (LIDC) [1]. They include lung nodules of multiple types, delineated in several trials by a group of radiologists and described in the form of a voxel-probability map M_{pr} . If all radiologists in all trials have indicated a voxel c to be a nodule, then $M_{pr}(c) = 100\%$; if none has done so, then $M_{pr}(c) = 0\%$, etc.

The method described above, named *M-OB*, has been evaluated. Two another approaches have also been tested and compared, both being parts of the *M-OB* method. The first one, called *Mask*, omits the main segmentation stage (fuzzy connectedness analysis), and passes the binary mask M_n as an input of the postprocessing stage. In the latter – *RFC-OB* – only a thresholded fuzzy connectivity scene serves as an input. In all three cases both, the pre- and post-processing stages, have been executed as described in sections 3 and 4.

The evaluation is based on the following measures:

- $P_{100\%}$ – percentage of voxels c assigned to a nodule with $M_{pr}(c) = 100\%$,
- $P_{50\%}$ – percentage of voxels c assigned to a nodule with $M_{pr}(c) \geq 50\%$,
- $N_0\%$ – percentage of false positives,

For each study, containing a single nodule, 5 pairs of seed points \mathbf{o}, \mathbf{b} have been indicated. For each pair the *Mask* algorithm has run once, the *RFC-OB* and *M-OB* algorithms – 5 times². Table 1 presents the results, along with the

² Due to the evolutionary algorithm used in the object seed points selection [3].

mean operation times t_{pr} , t_{op} , t_{po} for preprocessing, main, and postprocessing stages, respectively. The *M-OB* method provides the best results in M_{pr} -based factors, at 90 – 100% level of sensitivity and less than 0.5% of false positives. There are no significant differences in values obtained for particular types of nodules. However, the differences appear in operation times. As expected, the preprocessing stage requires the most time in case of pleura-connected nodules. The most time-consuming stage is the main one with a 3D fuzzy connectedness analysis, especially in case of large or low-intensity nodules.

Methodology described in this paper provides an effective segmentation of pulmonary nodules. Some specific difficult problems of the nodule appearance in CT studies are solved: connections between the nodule and pleura (by the preprocessing stage), vascularization (the postprocessing stage), and low mean intensity of a nodule region (the fuzzy connectedness analysis).

References

1. Armato, S., et al.: Lung Image Database Consortium: Developing a Resource for the Medical Imaging Research Community. *Radiology* 232, 739–748 (2004)
2. Badura, P.: Adaptive Thresholding in Fuzzy Approach to Segmentation of Cruciate Ligaments. In: *The Fifth International Workshop of Control and Information Technology, IWCIT 2006*, pp. 97–102 (2006)
3. Badura, P., Pietka, E.: Semi-Automatic Seed Point Selection in Fuzzy Connectedness Approach to Image Segmentation. *Advances in Soft Computing: Computer Recognition Systems 2(45)*, 679–686 (2007)
4. Badura, P.: Trojwymiarowa segmentacja guzow pluc z wykorzystaniem metod sztucznej inteligencji. PhD Thesis, Politechnika Slaska, Gliwice (2007)
5. Kostis, W., Reeves, A., Yankelevitz, D., Henschke, C.: Three-Dimensional Segmentation and Growth-Rate Estimation of Small Pulmonary Nodules in Helical CT Images. *IEEE Transactions on Medical Imaging* 22(10), 1259–1274 (2003)
6. Otsu, N.A.: Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* 9(1), 62–66 (1979)
7. Salambier, P., Oliveras, A., Garrido, L.: Antiextensive Connected Operators for Image and Sequence Processing. *IEEE Transactions on Image Processing* 7(4), 555–570 (1998)
8. Tadeusiewicz, R., Korohoda, P.: *Komputerowa analiza i przetwarzanie obrazów*. Wydawnictwo Fundacji Postępu Telekomunikacji, Kraków (1997)
9. Udupa, J., Saha, P.: Fuzzy Connectedness and Image Segmentation. *Proceedings of the IEEE* 91, 1649–1669 (2003)
10. Udupa, J., Samarasekera, S.: Fuzzy Connectedness and Object Definition: Theory, Algorithms, and Applications in Image Segmentation. *Graphical Models and Image Processing* 58(3), 246–261 (1996)
11. World Health Organization - Cancer (2007), <http://www.who.int/topics/cancer/en/>

Modeling and Simulation of Airway Tissues Stresses during Pulmonary Recruitment

Bożena Kuraszkiewicz

Institute of Biocybernetics and Biomedical Engineering PAS, Warsaw, Poland
bozena@ibib.waw.pl

Summary. In the present study the goal was to quantify the stresses acting locally on pulmonary epithelial cells in order to better understand the dynamic behavior of these cells. To quantify mechanotransduction responses, one must first understand the magnitude and distribution of stresses on the epithelial cells. It was investigated a two-dimensional, mathematical model of airway reopening, using a flow-driven semi-infinite bubble progressing through an airway as it clears a liquid occlusion was created. The flow in this system was highly viscous, and thus was governed by Stokes equations. This 2-D model was solved computationally using the boundary element method (BEM) in conjunction with lubrication approximations. Algebraic expressions were developed that could be used simply and accurately predict the fluid stress based upon the fluid viscosity, μ , channel height, H , cell size, A , and flow-rate, Q . From the solution, it was determined the stationary-state stresses acting on the epithelial cells. The results indicated that the magnitude of both the x - and y -stresses acting on the walls' cells were directly related to the cell protuberance topography and produced a complex stress field.

1 Introduction

A collapsed pulmonary system is one whose airways are blocked by a thin liquid film that obstructs airflow. If surface tension is large, the excessive pressure of mechanical ventilation is required to open the airways. Air and liquid occlusion flowing through the respiratory system exerts mechanical stresses on pulmonary cellular epithelium of a magnitude that damage the cells or modify their biological function. The tissues of the lung are delicate and highly sensitive to their mechanical environment. Abnormal physical forces, especially those associated with mechanical ventilation, potentially initiate or exacerbate lung injury [1]-[10]. During ventilation at low lung volumes and pressures, airway instability leads to repetitive collapse and reopening. Airway recruitment generates stresses (shear and normal) on the airway walls, potentially damaging airway tissues. The normal lung can tolerate repetitive collapse and reopening [9]. Pulmonary

surfactants reduce these stresses, but during steady reopening the surface tensions are higher than equilibrium. However, combined with insufficient or dysfunctional pulmonary surfactant, repetitive airway reopening produces severe lung injury [3, 4, 8]. Premature infants may have under-developed lungs that do not produce adequate surfactant that are required for normal lung function. This is known as Respiratory Distress Syndrome (RDS) [1, 2]. Mechanical ventilation necessary to sustain these infants can potentially cause severe damage to the sensitive lung tissues. RDS occurs in many infants born before the seventh month of gestation due to a pulmonary system, which is not yet fully mature. RDS results in exhaustion, inability to breath, and lung collapse and is the fourth leading cause of infant death today. With RDS and other obstructive pulmonary diseases, large pressures may be necessary to open collapsed airway for breathing.

Large mechanical stresses on pulmonary epithelial cells lining affected airways could cause damage or modify of their intended biological function [8, 11, 12]. All of these led as a motivation for the present study. To understand the physiological mechanisms of mechanotransduction within the pulmonary system, a mathematical, planar model for airway reopening, using a flow-driven semi-infinite bubble progressing through an airway as it clears a liquid occlusion was created in this study.

2 Model Description

To model this problem, the mechanics of semi-infinite gas bubble progression in a liquid-filled rigid-walled, axisymmetric tube under steady-state conditions was considered. Airway closure occurs when a liquid occlusion spans the airway and blocks airflow. This occlusion can be extensive or consists of a short meniscus. Both types of collapse will tend to occur simultaneously. However, local factors, such as the liquid volume and airway wall properties, will determine the predominant type of collapse. The airway was represented as a two-dimensional channel with corrugated walls to account for presence of cells. The model consisted of a horizontal channel, whose walls, separated by a mean distance ($2H$), were corrugated sinusoidally with amplitude (A) and wavelength (λ) (Fig. 1). The channel was occluded with a liquid occlusion of constant surface tension (γ), viscosity (μ) and density (ρ). The semi-infinite air bubble drove through an airway model with steady flow rate (Q). Adhesive properties of the airway lining fluid, such as the surface tension and viscosity, determined the airway reopening.

The system was described by the following dimensionless parameters:

- The ratio of viscous to surface tension forces/stresses
Ca—capillary number, that represents the dimensionless velocity,

$$\text{Ca} = (\mu Q)/(2H\gamma)$$

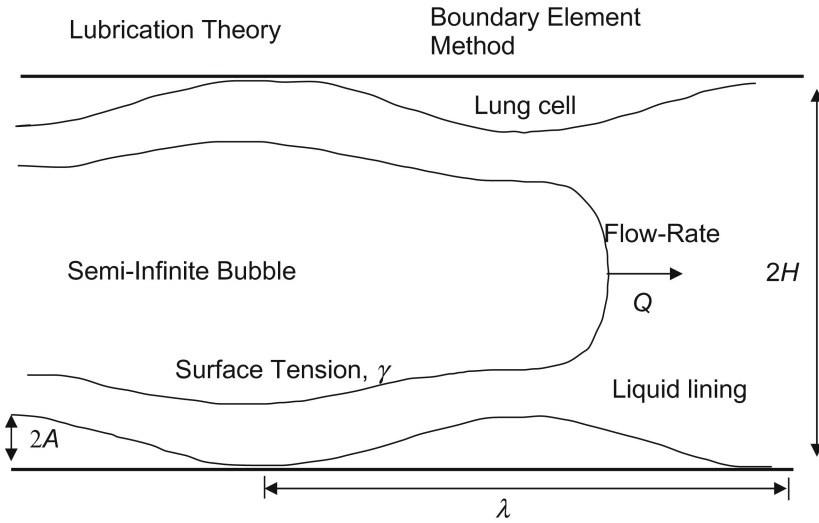


Fig. 1. An airway model with semi-infinite air bubble

- The ratio of cell average height (A) to half the average channel width (H),

$$\tau = A/H$$

Physiological values range is over $0.01 \leq \tau \leq 0.75$ from respiratory bronchioles to collapsed airways [7, 8, 12].

- The cell width (λ) to half the average distance separating the airway walls (H),

$$\Lambda = \lambda/H$$

Physiological values range is over $0.2 \leq \Lambda \leq 15$ from respiratory bronchioles to collapsed airways [7, 8, 12].

3 Method of Solution

The flow in this system was highly viscous, and thus was governed by Stokes equations [13]. It was assumed that inertia was negligible, based upon the Reynolds number $Re = UH/\nu \ll 1$, where U is a representative flow velocity and ν is the fluid kinematic viscosity. The viscous stresses were balanced by the fluid pressure gradient ($\nabla \cdot P$), so the hydrodynamic description was stated in the steady-state Stokes equations and continuity as:

$$\begin{aligned} \nabla \cdot P &= Ca \nabla^2 u \\ \nabla \cdot u &= 0 \end{aligned}$$

Because shear stress under Stokes flow is directly proportional to flow rate (Q), it is useful to represent the fluid mechanical interaction with the epithelial cells by dividing by flow rate (Q) [14, 15, 16]. This representation is beneficial for several reasons. First, it identifies the magnification of the mechanical influence on pulmonary tissue in an airway system with a fixed flow rate. Once flow rate is determined, it is simple to calculate the fluid mechanical impact on the cells. Most importantly, this representation depends only on physical constants of the system and the dimensionless flow-normalized response that is function only of dimensionless cells' average size $\tau = A/H$ [18, 19, 20]. This study assumed a constant surface tension (γ), therefore, neglected the influence of surfactant. Surfactant modifies the flow field by altering the local surface tension and mechanical stress balance at the interface.

This free-surface problem was solved using the boundary element method (BEM) in conjunction with lubrication approximations [14, 21]. Lubrication theory was used as a motivation for the development of a simple algebraic formula that could be used accurately to predict the mechanical influences over the range of different cell to channel height aspect ratios [22]-[26]. From the solution, it was determined the stationary-state stresses acting on the epithelial cells.

The study focused on the parameter physiological values:

$$\begin{aligned} \text{Ca} &= 0.01, \\ 0 \leq \tau &\leq 0.05, \\ A &= 2 \end{aligned}$$

and served to characterize the localized stress behavior at this cellular level.

4 Results

The results depict the dimensionless x - and y -stresses acting on the model's airway walls once the system has reached steady state. The data were collected when the tip of the moving bubbles was aligned with the top of a cell protrusion (in both cases).

The results demonstrate that a small cell protrusion can greatly amplify the stresses over what is experienced with a flat wall and introduce a complex, spatially dependent and temporally dependent stress field.

Fig. 2 presents the results for $\text{Ca} = 0.01$, $0 \leq \tau \leq 0.05$, $A = 2$.

Fig. 3 presents the results of data analysis and effects of ratio A/H on the magnitude of the x - and y -stresses.

These figures demonstrate the effect of an increase in dimensionless cell height ($\tau = A/H$) on the maximum change in the dimensionless x - and y -stresses acting on the models airway walls. Clearly, an extraordinary magnification of x - and y -stresses occurs with only a small amplitude cell corrugation.

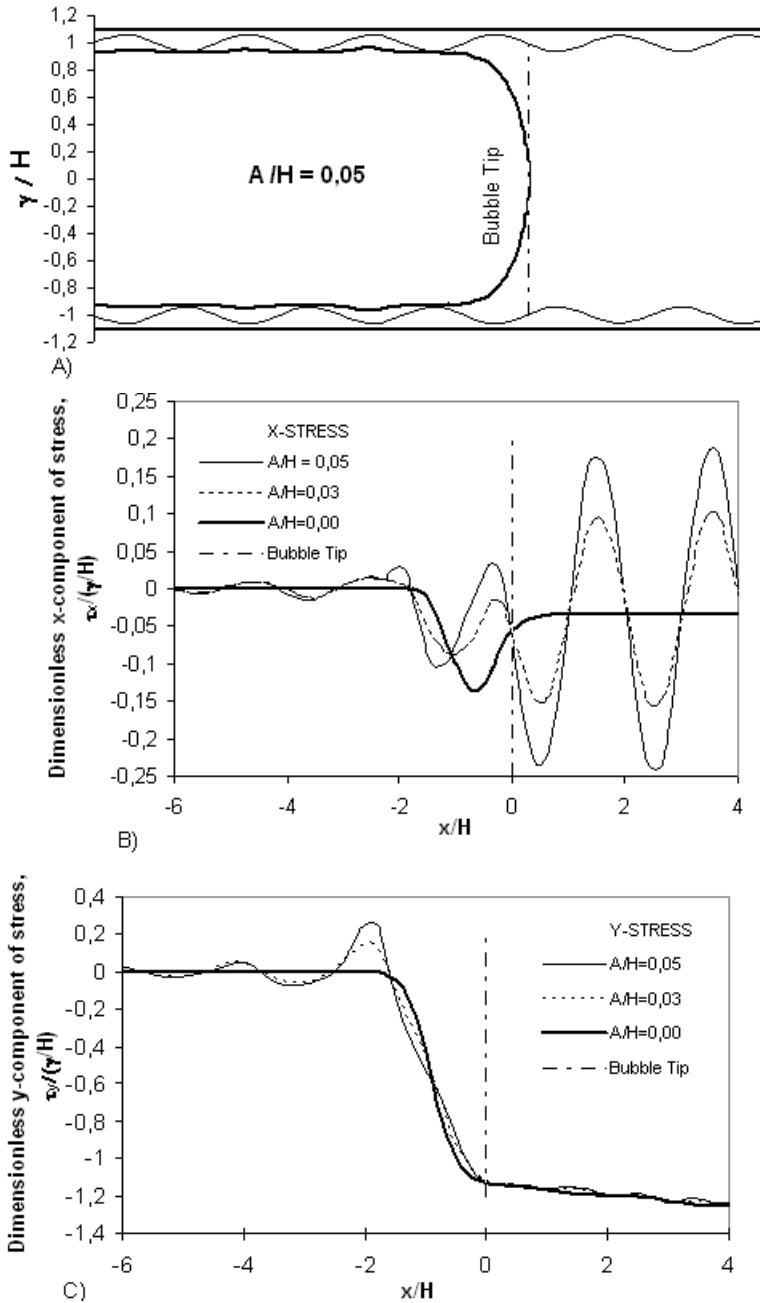
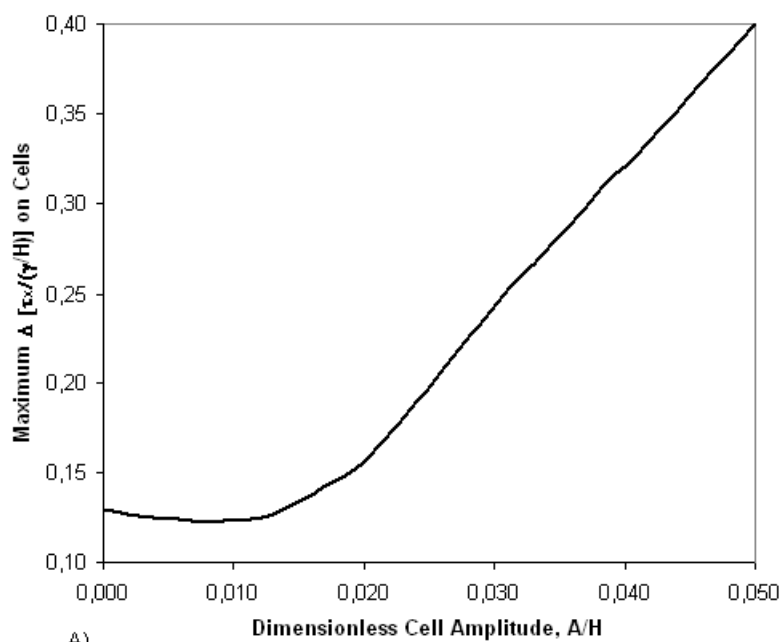
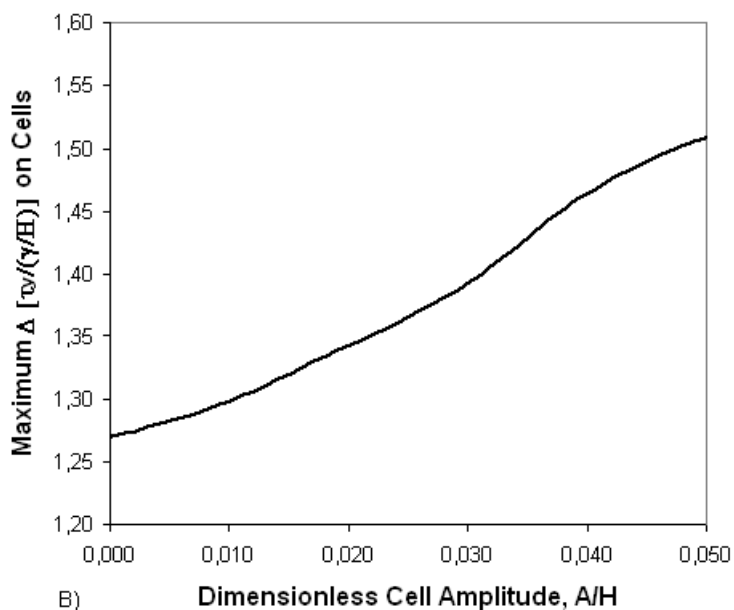


Fig. 2. The results for $Ca = 0.01$, $0 < \tau < 0.05$, $\Lambda = 2$. A) for domain $A/H = 0.05$; B) X-STRESS; C) Y-STRESS



A)



B)

Fig. 3. The results of data analysis. A) Effect of A/H on the magnitude of the x -stress; B) Effect of A/H on the magnitude of the y -stress.

5 Conclusions

In this study it has been modeled the flow field occurs in the vicinity of epithelial cells during airway reopening. The results indicate that the magnitude of both the x - and y -stresses acting on the walls' cells are directly related to the cell protuberance topography and produce a complex stress field.

The research demonstrates that varying values of fluid viscosity and surface tension result in perturbations of the stress field along the epithelium. The cell stresses are much larger than would be predicted using flat-walled assumptions. These large amplitude stresses may damage the cellular epithelium or modify its normal biological functioning.

Finally, it is important to recognize that this is a model study with many simplifications. Although it is expected that the predictions given here are accurate for rigid model of airway wall and epithelial cells, they ignore the influence of cell deformability and cell-to-cell interaction, which could clearly alter the behavior predicted in this study.

Acknowledgement. This work is supported by MNiSW Grant N51801732/1217.

References

1. Hyers, T.M., Fowler, A.A.: Adult respiratory distress syndrome: causes, morbidity, and mortality. *Annu Rev. Med.* 40, 431–446 (1998)
2. Murray, J.F., Matthay, M.A., Luce, J.M., Flick, M.R.: An expanded definition of the adult respiratory distress syndrome. *Am. Rev. Respir.* 139, 1065–1081 (1998)
3. Whitehead, T., Slutsky, A.S.: The pulmonary physician in critical care * 7: ventilator induced lung injury. *Thorax* 57, 635–642 (2002)
4. Dos Santos, C.C., Slutsky, A.S.: Invited review: mechanisms of ventilator-induced lung injury: a perspective. *J. Appl. Physiol.* 89, 1645–1655 (2000)
5. Burger, E.J., Macklem, P.: Airway closure demonstration by breathing 100% O₂ at low lung volumes and by N₂ washout. *J. Appl. Physiol.* 25, 139–148 (1968)
6. Pozrikidis, C.: Shear flow over a protuberance on a plane wall. *J. Eng. Math.* 31, 29–42 (1997)
7. Konstantopoulos, K., McIntire, L.V.: Cell adhesion in vascular biology. *J. Clin. Invest.* 98, 2661–2665 (1996)
8. Tschumperlin, D.J., Oswari, J., Margulies, A.S.: Deformation-induced injury of alveolar epithelial cells. Effect of frequency, duration, and amplitude. *Am. J. Respir. Crit Care. Med.* 162, 357–362 (2000)
9. Taskar, V., John, J., Evander, E., Wollmer, P., Robertson, B., Jonson, B.: Healthy lung tolerates repetitive collapse and reexpansion. *Acta Anaesthesiol Scand* 39, 370–376 (1995)
10. Mead, J., Takishima, T., Leith, D.: Stress distribution in lungs: a model of pulmonary elasticity. *J. Appl. Physiol.* 28, 596–600 (1970)
11. Brooks, S.B., Tozeren, A.: Flow past an array of cells that are adherent to the bottom plate of a flow channel. *Computers Fluids* 25, 741–757 (1996)
12. Characklis, W.G., Marshall, K.C.: *Biofilms: A basis for an interdisciplinary approach.* John Wiley and Sons, New York (1990)

13. Higdon, J.J.L.: Stokes flow in arbitrary two-dimensional domains: shear flow over ridges and cavities. *J Fluid Mech.* 159, 195–226 (1985)
14. Halpern, D., Gaver, D.P.: Boundary element analysis of the time-dependent motion of a semi-infinite bubble in a channel. *J. Comput. Phys.* 115, 366–375 (1994)
15. Ladyzenskaya, O.A.: The mathematical theory of viscous incompressible flow. Gordon and Breach, New York (1963)
16. Leal, L.G.: Laminar flow and convective transport processes: Scaling principles and asymptotic analysis, Butterworth–Heinemann, Boston (1992)
17. Oliver, L.A., Truskey, G.A.: A numerical analysis of forces exerted on spreading cells in a parallel plate flow chamber assay. *Biotechnol Bioeng* 42, 963–973 (1993)
18. Fung, Y.C.: *Biodynamics: Circulation*. Springer, New York (1984)
19. Davies, P.F.: Flow-mediated endothelial mechanotransduction. *Physiol. Rev.* 75, 519–560 (1995)
20. Dillon, R.L., Fauci, L., Fogelson, A., Gaver, D.P.: Modeling biofilm processes using immersed boundary method. *J. Comput. Phys.* 129, 57–73 (1996)
21. Brebbia, C.A., Dominguez, J.: *Boundary elements—an introductory course*, Southampton, England. Computational Mechanics (1989)
22. Dillon, R.L., Fauci, L., Gaver, D.P.: A microscale model of bacterial swimming, chemotaxis and substrate transport. *J. Theor. Biol.* 177, 325–340 (1995)
23. Ingber, D.E.: Tensegrity, the architectural basis of cellular mechanotransduction. *Annu Rev. Physiol.* 59, 575–599 (1997)
24. Gaver, D.P., Halpern, D., Jensen, O.E., Grotberg, J.B.: The steady motion of a semi-infinite bubble through a flexible-walled channel. *J. Fluid Mech.* 319, 25–65 (1996)
25. Ghadiali, S.N., Gaver, D.P.: The influence of non-equilibrium surfactant dynamics on the flow of a semi-infinite bubble in a rigid cylindrical tube. *J. Fluid Mech.* 478, 165–196 (2003)
26. Suki, B., Barabasi, A.L., Hantos, Z., Petak, F., Stanley, H.E.: Avalanches and power-law behaviour in lung inflation. *Nature* 368, 615–618 (1994)

Compression of Bronchoscopy Video: Coding Usefulness and Efficiency Assessment

Artur Przelaskowski and Rafal Jozwiak

Institute of Radioelectronics Warsaw University of Technology, Nowowiejska 15/19,
00-665 Warsaw, Poland

arturp@ire.pw.edu.pl, rjozwiak@ire.pw.edu.pl

Summary. This paper aims at quality assessment and control of bronchoscopy video data used for diagnosis and managed in medical information systems. Application of different video coding/decoding methods according to MPEG and JPEG family standards were considered for effective video data storage and communication. Enhanced image sequences preview, analysis and interpretation according to radiological procedures is possible because of improved, flexible, hierarchically ordered data representation. Useful coding or transcoding algorithms were studied, experimentally optimized and verified according to quality preserving criteria based on objective numerical measures, subjectively controlled. Compression schemes suitable for bronchoscopy video were concluded.

1 Introduction

Medical video examination has become a very useful way of diagnosis last years. Together with modern 3D visualization methods, advanced medical video techniques has shown a marked improvement in diagnosis and treatment efficiency but also in the fields of medical education, training and presentation. Medical image sequences come mainly from endoscopic examinations, i.e. gastroscopy, colonoscopy, bronchoscopy, enteroscopy, laparoscopy, cystoscopy etc. Advances in video technology allow visual inspection for diagnosis or treatment of the inside of the human body without or with very small scars. During the endoscopic procedure, the tiny video camera records a video signal of the interior of the human organ, which is displayed on a monitor for real-time analysis by the physician [1]. Other medical imaging sources of image sequences are angiography, ultrasound examinations: sequential B-mode, color Doppler, power Doppler, ultrasound contrast imaging, dynamical tomography (CT or MR).

Diagnosis based on medical video requires effective and adjustable representation of data stored in huge medical image databases, communicated across integrated medical information systems, distributed and interpreted in real time through band-limited telecommunications channels for teleradiology purposes etc. Efficient data management and necessity of adaptation to the requirements of different system elements (see Fig. 1) require flexible and quality controlled coding/transcoding methods which preserve full diagnostic information and useful data stream features adapted to receiver characteristics. Among employable

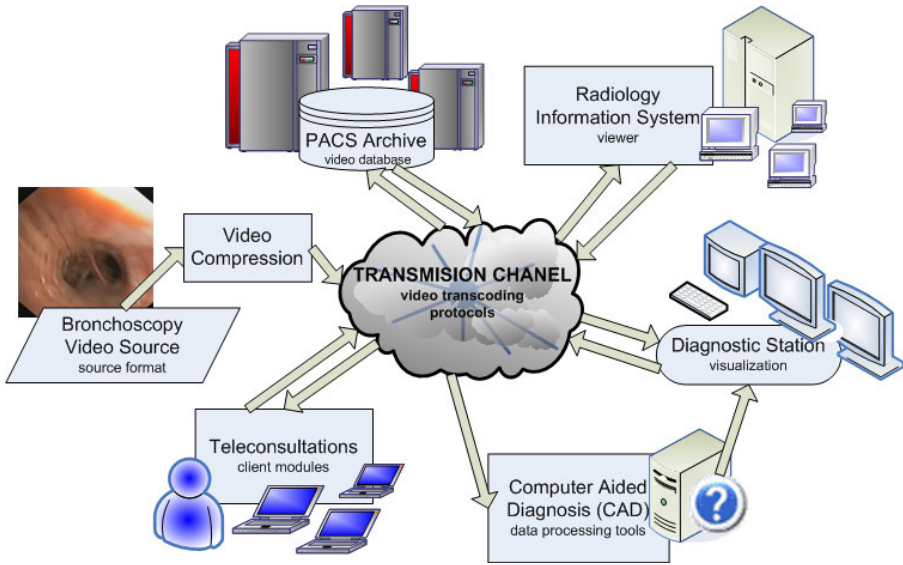


Fig. 1. Diagram of integrated diagnostic system with bronchoscopy video data flow. Efficient source data compression or transcoding due to receiving equipment parameters and application conditions is required.

and desired codec enhancements, scrolling with both automatic redundant (diagnostically unimportant) information reduction and specific (diagnostically important) information selection, diagnostic features scalability and progression, indexing for fast content-based image retrieval and data transmission error resilience can be mentioned.

The most challenging task of video encoding is to balance between preserving accurate quality and achieving satisfying representative bit rate. Low bit rate of video data stream is often required due to the limitation of storage or transmission processes [2]. On the one hand, for the purpose of reliable diagnosis it is necessary to keep medical video data "diagnostically lossless" or sometimes even "visually lossless". Loss of important details, smoothed and less perceptible edges or introduced noticeable artifacts are strictly forbidden. Determining formal criteria of interpreted examination quality, selection of medical video coding standards and defining of acceptable compression bit rates is nowadays one of the most important development task for DICOM¹ standardization process². In details, important issues are as follows:

- estimation of diagnostically lossless bit rates for designed coding/transcoding methods and characteristics of video-based examinations;
- selection and adjusting of the coding methods to specific application requirements taking into consideration video quality indicators, range of acceptable

¹ Digital imaging and communications in medicine.

² <http://www.dclunie.com/dicom-status/status.html>

bit rates, useful scales or resolution, information progressive mode, contrast sensitivity of preferred region of interests (ROI) etc.;

- selection of the transcoding methods due to required video data properties, alternating transmission channel specification or customer expectations;
- adaptation of reliable video quality assessment methods, numerical or subjective, to control or verify video quality, to optimize codecs or transcoders, to optimize interactively parameters of communicated data streams etc.

Specific medical imaging modality used to verify usefulness and optimize video coding methods is bronchoscopy. Bronchoscopic examinations demonstrate many common features with natural video sequences, e.g. general image features and natural content perception, color space, textural features, data dynamics, dominant objects properties, movement characteristics. In this paper important aspects of medical video compression, especially a choice and the optimization of coding/transcoding methods were considered. Moreover, numerical metrics were analyzed to find the most useful measures of video diagnostic accuracy for compression purposes. Selected codecs and quality measures of coded bronchoscopy video were verified experimentally.

2 Methods

For the sake of interpretation process, which quite often depends on single freeze frame analysis, it is very important to preserve high quality for each compressed frame. Therefore making any irreversible processing requires thoughtful exercise and accuracy assessment. Some standards, like M-JPEG2000 allow for numerically lossless video compression at the cost of quite low compression ratio. Another, quit safe and acceptable compression that allows achieving higher ratios assumes on weak quantization providing visually lossless video representation: sensible, acceptable compression limit depends on observer ability of distortion perception in the reconstructed video. Additional video data size reduction is possible with diagnostically important ROI selection, where data are compressed losslessly or near losslessly. Diagnostically unimportant regions are highly distorted to decrease bit rate without diagnostically noticeable consequences.

2.1 Video Compression

Because of the results of DICOM standardization efforts and the last studies reported in literature one can state that among many modern video codecs MPEG-2, MPEG-4 (according to part II) and H.264 (advanced video codec, part 10 of MPEG-4) seem to be the most suitable for medical video compression [2, 3, 4, 5]. Besides common popularity and high efficiency applied to natural video, these codecs are characterized by number of useful features crucial for the purpose of medical video compression and were or are considered to be included to DICOM standard.

Recent successes of H.264 applications for medical and non-medical video compression and notified improvements of H.264 codec implementations make

this paradigm of video compression preferable for most purposes. However, other less sophisticated video compression standards like DV (digital video) and M-JPEG2000 (JPEG2000, part 3 – *motion*) were sometimes preferred for medical video [6, 7]. Thus design of motion estimation scheme for inter- and intraframe mode based on subpixel object characteristics with well fitted quantization procedure and context-based binary arithmetic encoding adapted to progressive and scalable data passes, ROI progression with lossless compression mode, error resilience and rate distortion optimization of ordered data stream seem to be crucial for high video compression effectiveness. Inter alia, H.264 enables six different macroblock coding modes inter (for block size of 16x16, 16x8, 8x16, 8x8) and intra (for block size of 16x16 and 4x4) for motion estimation or skipping with quarter-pixel precision for motion compensation selected according to Lagrangian rate-distortion procedure. The H.264 standard determines four predefined coding profiles: *baseline* (I/P-frames support, some enhanced error resilience tools (FMO, ASO, and RS), supports only progressive video and Context-Adaptive Variable Length Coding (CAVLC)), *main* (I/P/B-frames support, progressive and interlaced video coding, CAVLC or Context-Adaptive Binary Arithmetic Coding (CABAC)), *extended* (I/P/B/SP/SI-frames support, some enhanced error resilience tools, but like *baseline* supports only progressive video and CAVLC) and *high* (adds to *main* adaptive transform block size (8x8 intra prediction), quantization scaling matrices, lossless video coding, more yuv formats, etc.) [8]. Moreover, the standard defines multiple motion estimation modes through selection of comparison function used for fullpixel motion estimation: diamond, hexagon, multi-hexagon, and exhaustive motion search (with definable search range). Optimized realizations of H.264 codec fitted to specific requirements of medical examination video is still open and challenging question.

During video data stream transcoding, first of all problems regarding a correction and possible reduction of input lossy compression distortions completed with adjusted new distortions of quantization procedure, adaptive full and flexible scalability and progression of output video stream are in the limelight. The limitation of quantization distortion accumulation as well as additional usage of perception improvement methods which are able to adapt easily to observer characteristics, diagnostic features of image data and technological conditions of visualization process are extremely important. Therefore employment of reliable video quality metrics is necessary and fundamental for the reliability of data representation optimization and medical video information management procedures. Initial experiment with adjusting H.264 compression to transcoding challenges resulted in satisfying converting efficiency [9].

2.2 Numerical Video Quality Assessment

The medical video content is quite varied and characterized by many features like diversified background, variable mean brightest level, local and global contrast, variable noise level, movement and enhancement of diagnostic structures over-reflected areas (frames) etc. Additionally, in medical video quality assessment, inter- and intra-frame semantic image content plays an important role.

Beside subjective quality assessment, i.e. ratings with properly established scales (like MOS³ widely known from ITU norms) or pathology detection tests with statistical analysis (methods based on ROC⁴), reliable numerical metrics are first and foremost searched. Such computationally objective measures should highly correlate to subjective radiologist interpretation, additionally objectified by "gold standard" patterns. It is crucial to adapt numerical quality indicators to the requirements of specific applications taking into consideration expert operating characteristics and diagnostically important descriptors of analyzed content.

Considering commonly used, easy to implement numerical measures, e.g. mean square error (MSE), the drawback is generally poor correlation to the quality of diagnostic content. Enhanced measures based on simple human visual system (HVS), error weighting with contrast sensitivity function (CSF), visual progressive weighting (VIP) or additional masking or normalization methods dependent on information perception have also limited quality assessment efficiency [10]. Among quality measures which seem to be more suitable for radiological purposes and can be easily adapted for video quality assessment are vector metrics reduced to scalar quality indicators in subjectively-driven optimization procedures. An important role in this group of measures play Picture Quality Scale (PQS) and Hybrid Vector Measure (HVM) [11] and additionally Structure Similarity Index Measure (SSIM) [12]. Alternatively, Video Quality Metrics (VQM) with more sophisticated, DCT-based HVM modeling occurred useful for codecs verification [13].

Nowadays methods based on manual or automatic region of interest (ROI) selection, tracking and quality assessment in these regions by means of measures which take into consideration different types of distortions and reliable HVS models, seem to be the most useful. Unfortunately we suffer from a lack of reliable temporal models of diagnostic content sequences, which can be used for numeric metrics construction or objectivization of subjective ratings. Visual perception effects like temporal masking, i.e. an elevation of visibility thresholds due to temporal discontinuities in intensity, adjusting the sensitivity of the visual system in response to the prevalent stimulation patterns, low-pass sustained and high-pass transient mechanisms etc, are extended to diagnostic analysis and interpretation models of the video examinations. Expert quality of experience models lead to the most reliable quality measures. Between the propositions more or less approximating such quality assessment paradigm are:

- SSIM-video where structural distortion is used for an estimate of perceived visual distortion; the quality was estimated in the local image regions, frames and in the sequence with motion level estimates and luminance frame weighting for content perception sensitivity modeling [14];
- perceptual distortion metric (PDM) that is based on a contrast gain control model of the human visual system that incorporates spatial and temporal aspects of vision as well as color perception [15];
- hybrid vector measure for video (HVM-V) especially adapted to medical imaging applications [9].

³ Mean opinion score.

⁴ Receiver operating characteristics.

HVM-V was adjusted adaptively to examination specification (modality conditions, diagnostic content) and uses inter- and intraframe perception models of video data. HVM-V is composed of 6 scalar metrics which describe several kinds of distortion according to diversified visual perception effects: local or global distortions characteristics, random error measure (normalized error energy weighted by perception model) and local structural error measures, i.e. local spatial correlation and psycho-physical effects which affects the perception of errors in the vicinity of high contrast transitions. HVM-V is distinguished by universality and flexibility. Factors are match semiadaptationally, based on mean distortion value for each measure and subjective observer preferences at preliminary stage.

3 Experiments

Four bronchoscopy video test sets were used in the experiments (see Fig. 2 and Tab. 1). A group of selected video codecs which seem to be preferable for

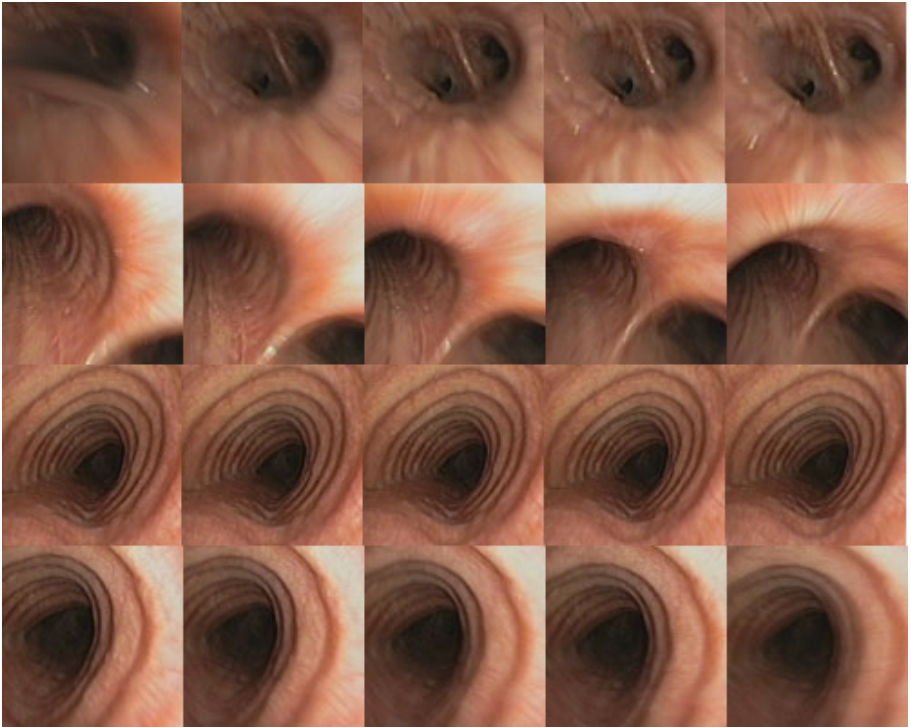


Fig. 2. Four bronchoscopy video test sequences. First two (top ones) are distinguished by fast camera movement, variability of characteristic and reflections while on another two sequences (bottom ones) we can observe slow, uniform camera movement and stable kind of environment features.

Table 1. The characteristics of bronchoscopy video sets used in compression experiments

Sequence	Numbers of frames	Frame rate [fps]	Resolution	Duration [s]	Size [KB]
broncho1	175	25	256x256	7	33600
broncho2	200	25	256x256	8	38400
broncho3	185	25	256x256	7.4	35520
broncho4	189	25	256x256	7.56	36288

Table 2. The experimental results of bronchoscopy video compression. Default options of standard codecs were used. The following implementations of standard codecs were used: OpenJPEG (www.openjpeg.org) for M-JPEG2000 and FFmpeg (ffmpeg.mplayerhq.hu) for others.

Sequence	Compression size ~133 KB	Objective numerical quality assessment			
		PSNR	SSIM	VQM	HVM-V
broncho1	M-JPEG2000	41.02	0.957	0.745	95.58
	MPEG-2	25.91	0.776	3.775	400.86
	MPEG-4	39.46	0.943	0.855	127.47
	H.264	41.72	0.96	0.714	92.99
broncho2	M-JPEG2000	38.21	0.931	0.979	131.45
	MPEG-2	26.16	0.742	3.581	399.06
	MPEG-4	37.32	0.918	1.039	160.5
	H.264	39.54	0.942	0.855	114.01
broncho3	M-JPEG2000	33.91	0.847	1.59	224.4
	MPEG-2	30.6	0.756	2.32	344.97
	MPEG-4	37.47	0.93	1.018	122.86
	H.264	37.53	0.936	1.099	125.73
broncho4	M-JPEG2000	34.65	0.868	1.459	191.48
	MPEG-2	29.2	0.737	2.695	355.09
	MPEG-4	36.84	0.908	1.121	141.79
	H.264	37.55	0.925	1.069	125.79
<i>broncho mean</i>	M-JPEG2000	36.95	0.901	1.193	160.73
	MPEG-2	27.97	0.753	3.093	375.00
	MPEG-4	37.77	0.925	1.008	138.16
	H.264	39.09	0.941	0.934	114.63

medical video compression were selected and verified. Codecs were optimized through selection of motion estimation modes, adjusting quantization and coding parameters. To assess quality of compressed video, selected numerical measures and subjective test of acceptable bit rate estimates for visually lossless compression were used. The results of the experiments were presented in Tables 2, 3, 4.

We tried to determine acceptable visually lossless compression rate. For two different bronchoscopy video sequences (with various visual features): broncho1

Table 3. The experimental results of bronchoscopy video compression. H.264 optimization according to typical profiles was underlined – *baseline* and *main* profiles (with default x264 encoder options), and *high* profile (with adaptive DCT and all possible macroblock analysis sizes) were used. The best results were bolded.

Sequence	Compression size ~133 KB	Objective numerical quality assessment			
		PSNR	SSIM	VQM	HVM-V
<i>broncho mean</i>	H.264 base	38.24	0.932	1.020	128.02
	H.264 main	39.09	0.941	0.934	114.63
	H.264 high	39.47	0.945	0.917	112.89

Table 4. The experimental results of bronchoscopy video compression. H.264 optimization according to searching procedure for motion estimation was made – all methods (diamond, hexagon, multi-hexagon, and exhaustive) with two search ranges: 16 and 32 for exhaustive motion mode were tested. Mean values for 4 test videos were presented.

Sequence	Compression size ~133 KB	Objective numerical quality assessment			
		PSNR	SSIM	VQM	HVM-V
<i>broncho mean</i>	H.264 hex	39.47	0.945	0.917	112.89
	H.264 dia	39.45	0.945	0.918	113.10
	H.264 mhex	39.50	0.945	0.914	112.65
	H.264 exh	39.51	0.945	0.912	113.01
	H.264 exh32	39.52	0.945	0.909	112.96

and broncho3 subjective video quality assessment procedure based on modification of SCACJ⁵ [16] was carried out. Two image sequences were played simultaneously (original and compressed with H.264 optimized to different rates), and observers decided whether videos are identically or not. The acceptable visually lossless compression rates were determined as mean rate value for which experts stopped to perceive meaningful difference between both video sequences. Four experts in image and video processing participating this experiment. The resulted acceptable bit rates for visually lossless compression based on H.264 were estimated as follows: 0.84 bpp (172 kbps) for dynamic video (broncho1) and 0.53 bpp (108 kbps) for static video (broncho3).

4 Conclusions

Lossy compression of bronchoscopy video seem to be acceptable for diagnosis because of visually lossless reconstruction of interpreted video data in estimated bit rate range. Compression bit rates up to 172 kbps can be safely used for storage and communication purposes. H.264 compression standard gives the best quality

⁵ Stimulus comparison adjectival categorical judgment.

of compressed video according to all used quality metrics, especially to reliable vector accuracy measures often used for verification of medical video and image compression. Optimization of H.264 codec increased compression efficiency with noticeable subjective enhancement especially for *high* coding profile. Alternative M-JPEG2000 codec seem to be really useful for bronchoscopy video compression with the effectiveness comparable to H.264 for dynamic video with fast camera movement across variable diagnostic content.

References

1. Oha, J., Hwangb, S., Leeb, J., Tavanapongc, W., Wongc, J., de Groend, P.C.: Informative frame classification for endoscopy video. *Medical Image Analysis* 11, 110–127 (2007)
2. Paul, M., Sorwar, G.: Encoding and decoding techniques for medical video signal transmission and viewing. In: *IEEE Conf Comput Inform Science (IEEE-ICIS-2007)*, pp. 750–756 (2007)
3. Yu, H., Lin, Z., Pan, F.: Applications and improvement of H.264 in medical video compression. *IEEE Tran. Circ. Sys. I, Spec issue Biomedical Circuits and Systems: A New Wave of Technology* 52(12), 2707–2716 (2005)
4. Yu, H., Lin, Z., Tan, R.S., Le, T.T., Ghista, D.N.: H.264 standard for left ventricle video compression and telediagnosis. In: *Proc IASTED Telehealth*, p. 564 (2007)
5. Frankewitsch, T., Söhnlein, S., Müller, M., Prokosch, H.-U.: Computed quality assessment of MPEG4-compressed DICOM video data. *Stud. Health Technol. Inform.* 116, 447–452 (2005)
6. Fossel, S., Fottinger, G., Mohr, J.: Motion JPEG2000 for high quality video systems. *IEEE Tran. Consum. Elect.* 49(4), 787–791 (2003)
7. Bezan, S., Shirani, S.: RD optimized, adaptive, error-resilient transmission of MJPEG2000-coded video over multiple time-varying channels. *EURASIP J. Appl. Sig. Proces.* 1, 1–13 (2006)
8. Richardson, I.: *H.264 and MPEG-4 video compression: video coding for next-generation multimedia*. John Wiley & Sons, Chichester (2003)
9. Przelaskowski, A., Jozwiak, R.: Kompresja i transkodowanie medycznego wideo: metody kontroli jakości danych (Coding and transcoding of medical video: data quality control). *Elektronika* (in press, 2008)
10. Przelaskowski, A., Jozwiak, R., Krzyzewski, T., Wroblewska, A.: The ordering of diagnostic information in encoded medical images: accuracy progression. *Optoelectronics Review* 16(1), 49–59 (2008)
11. Przelaskowski, A.: Falkowe metody kompresji danych obrazowych. ch. 5. In: *Oficyna Wydawnicza Politechniki Warszawskiej* (in polish, 2002)
12. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Tran. Image Proc.* 13(4), 600–612 (2004)
13. Xiao, F.: DCT-based video quality evaluation, Technical report, Final Project for EE392J, Stanford University (2000)
14. Wang, Z., Lu, L., Bovik, A.C.: Video quality assessment based on structural distortion measurement. *Signal Proc: Image Comm.* 19(2), 121–132 (2004)
15. Winkler, S.: *Digital video quality*. Wiley & Sons, Chichester (2005)
16. Recommendation ITU-R BT.500-7, Methodology for the subjective assessment of the quality of television pictures. ITU-R, Geneva (1974-1997)

Fuzzy Rule-Based System for the Diagnosis of Laryngeal Pathology Based on Contact Endoscopy Images

Wojciech Tarnawski and Jacek Cichosz

Chair of Systems and Computer Networks, Wrocław University of Technology
wojciech.tarnawski@pwr.wroc.pl, jacek.cichosz@pwr.wroc.pl

Summary. In this paper the fuzzy rule-based system for nuclei classification is presented. Firstly, in order to receive proper partition of objects (nuclei) the definition of features which can be used for diagnosis of laryngeal pathology based on contact endoscopy images is described. After the feature selection the fuzzy clustering process is realized. It creates the set of training input-output data pairs which later are used for generation of fuzzy rules by means of the method called learning from examples.

1 Introduction

The rule-based system symbolic processing could be the best platform to explain physician knowledge in medical diagnosing. These rules can be extracted from expert/physician knowledge or learned from examples. In daily life diagnosis, quite often physicians use some sort of fuzzy rules to interpret what they perceive to give the correct diagnosis. In this paper the fuzzy rule system for the quantification and classification of cell images derived from the contact endoscopy is presented. Contact endoscopy is a technique [1, 9] used to obtain detailed magnified images of living epithelium using a modified glass rod lens endoscope placed (hence: contact endoscopy) on the surface of the tissue. The input for the proposed system is the segmented contact endoscopy image, where the segmentation approach is presented by author in [7]. The system output can be used in the computer-aided intra-operative cancer diagnosis of the larynx. The main problem with diagnostic interpretation for cancer diagnosis with the contact endoscopy images is connected with the intuitive description of the image objects attributes such as: "rather bigger size of the cells nuclei", "highly deformed shape of the nuclei" and "high density of cell nuclei" or "the cells nuclei are grouped very closely". All these attributes defy precise description of the image objects such as: object size in pixels, value of the chosen shape coefficient etc., and they are best modeled by fuzzy sets. This paper presents a straightforward algorithm [8] called learning from examples for learning fuzzy rules from numerical data. In this approach a set of desired input-output data pairs (the training data) is required. But in our case only input values (object features: object field, object shape coefficient etc.) from training data are available; outputs (class label for every object) are generated during feature space clustering. Feature space

clustering is realized with Gustaffson-Kessel fuzzy clustering approach [4] where the clustering results from sample images taken for the system learning should be accepted by an expert/physician. The task here is to generate a set of fuzzy rules from the received training data and use them for further description and classification of the contact endoscopy images. Outputs of the proposed fuzzy rule-based system in the near future will be confirmed by histopathology defined in our research as the "golden standard".

The structure of this paper is organized as follows. In the next section the feature selection problem is described. The following section concerns fuzzy clustering approach used for supervised object labeling. Next section is composed of four subsections describing steps for fuzzy rules generation by learning from examples. The following section describes a defuzzification method used for mapping between inputs and the output. In the last section we present the system results for contact endoscopy images and few conceptions for future work.

2 Feature Selection

All tissues are composed of cells including nucleus, and the critical examinations of this biological material is important to an understanding of biological processes, both normal and abnormal (pathological). Abnormal nuclei by are described by an expert as: "rather bigger size of the cells nuclei", "highly deformed shape of the nuclei" and "high density of cell nuclei" or "the cells nuclei are grouped very closely". Therefore the computer analysis should give evaluation of the cell structures: i.e. the presence of nuclei, their size and color or intensity, shape analysis of the nuclear and the nucleus/cytoplasm ratio or distances between nucleus. The single, segmented i -th object on the segmented image is described by the input vector $x^{(i)} = [x_1^{(i)}, x_2^{(i)}, \dots, x_q^{(i)}]^T$, where q is the dimension of the input vectors. In this paper the concept of an *object* identifies single nucleus described by three features ($q = 3$). The first feature $x_1^{(i)}$ is termed *object field* and is expressed by the number of pixels composing i -th object. The second $x_2^{(i)}$ is related to the *shape of the object* and is expressed by shape coefficient proposed by Blair and Bliss and is defined as:

$$x_2^{(i)} = \frac{x_1^{(i)}}{\sqrt{2\pi \sum_k r_k^2}}, \quad (1)$$

where r_k is the distance between selected k -th pixel belonging to the i -th object and its centroid. The third feature $x_3^{(i)}$ is a local measure of *density of objects* and indirectly describes *distances between neighboring objects contours* (nuclei). Living epithelium cells touch each other and this feature indirectly expresses nucleus/cytoplasm ratio because distance between nuclei is related to the size of cytoplasm surrounding nuclei. The novel idea presented in this paper is used to formalize linguistics defined as "high density of cell nuclei" or "the cells nuclei are grouped very closely", is based on the partition of the image space using multiple grids of different size, also called multi-resolution grids. The proposed

method assumes that analyzed image space is quantized by K grids composed from rectangles. The size of the single mesh in the initial grid ($k = 0$) is $(2^p)_{k=0} \times (2^p)_{k=0}$, where p is the power parameter defined by the user. Every single mesh in the k -th grid, where $k = 1, 2, \dots, K - 1$ is divided into four, every of size $(2^{p-k})_k \times (2^{p-k})_k$. The size of a single mesh is expressed by power of two because it makes possibility for further fast computation by using bit-shift operations. The algorithm counts, through out the grids, the values of density function $d_k(m, n)$ in the k -th grid for every pixel of the image space localized at (m, n) according to the formula:

$$d_k(m, n) = \sum_{(m,n) \in V_{(m_0,n_0)}^k} I_{object}(m, n) / V_{(m_0,n_0)}^k, \tag{2}$$

where

$$I_{object}(m, n) = \begin{cases} 1 & \text{if pixel } (m, n) \text{ belongs to an object,} \\ 0 & \text{if pixel } (m, n) \text{ belongs to the background.} \end{cases} \tag{3}$$

$V_{(m_0,n_0)}^k$ denotes a single mesh in the k -th grid with its centroid located at (m_0, n_0) , and $|V_{(m_0,n_0)}^k|$ denotes the size of this mesh expressed in pixels. The multi-resolution conception is based on the fact that grids of different sizes take into consideration the contribution of density distribution from different points of view and better approximates distribution of the nuclei in the analyzed tissue. Therefore, the final definition of the third feature $x_3^{(i)}$ for the i -th object is described by the mean of the interpolated and weighted superposition of the multi-resolution values D of the density function at different resolutions calculated for every object pixel:

$$x_3^{(i)} = \sum D(m, n) / x_1^{(i)}, \quad \text{if } I_{object}(m, n) = 1, \tag{4}$$

where $D(m, n) = \sum_{k=0}^{K-1} w_k \cdot \bar{d}_k(m, n)$ and w_k are values (weights) of the Gaussian distribution with standard deviation defined by the user which are calculated in the initial state of the algorithm. It gradually weights the values of the interpolated density function in the following grids by ascribing the highest values for $k = K - 1$ -th grid i.e. with the smallest size of the mesh; $\bar{d}_k(m, n)$ denotes the interpolated density function in the k -th grid calculated according to the formula:

$$\bar{d}_k(m, n) = \sum_{s=1}^S [d_k(m_s, n_s) \cdot P_s] / |V_{(m_0,n_0)}^k|, \tag{5}$$

where (m_s, n_s) are centroids of the nearest meshes ($S = 4$) related to the point (m, n) and P_s are the linear interpolation coefficients expressed in pixels. Finally, this value is scaled using the quasi-sigmoidal transfer function.

In this way, the three features define 3-dimensional feature space used for further object classification where the single i -th object on the segmented image is described by the input vector: $x^{(i)} = [x_1^{(i)}, x_2^{(i)}, x_3^{(i)}]^T$.

3 Feature Space Clustering

Feature space clustering is realized with Gustaffson-Kessel fuzzy clustering approach [4], where the clustering results at few images chosen for system learning by an expert/physician should be accepted by the same expert. Therefore this stage could be called supervised object labeling producing a set of desired input-output data pairs (the training data):

$$([x_1^{(1)}, x_2^{(1)}, x_3^{(1)}]^T; O^{(1)}), ([x_1^{(2)}, x_2^{(2)}, x_3^{(2)}]^T; O^{(2)}), \dots \quad (6)$$

where $O^{(i)}$ is the output for the i -th object on the segmented image denoting the class number. This value is determined on the basis of the maximum membership value principle (the winning cluster, depicted the class number in this case, has the maximum membership function value for the i -th object). In our approach the expert's acceptance can be realized in two ways: 1) by defining the number of classes visible on the analyzed image or 2) by accepting the number of classes proposed by the fuzzy clustering algorithm based on some cluster validity measures [5, 6] with clustering results on the image. In our approach we have used the following cluster validity measures defined for the F objects and C clusters (classes); u_{ij} denotes membership of the i -th object to the j -th cluster (class) and $v^{(j)}$ denotes prototype of the j -th cluster:

- 1) *Partition coefficient*: $PK = \frac{1}{F} \sum_{j=1}^C \sum_{i=1}^F u_{ij}^2$,
- 2) *Separation coefficient*: $S = \sum_{i=1}^F \sum_{j=1}^C u_{ij}^\eta (\|x^{(i)} - v^{(j)}\|^2 - \|\sum_{i=1}^F / F - v^{(j)}\|^2)$, where η denotes the exponent factor called fuzzifier [2] defined also for the clustering process, and
- 3) *Partition coefficient*: $CS = \sum_{i=1}^F \sum_{j=1}^C u_{ij}^2 \|v^{(j)} - x^{(i)}\| / (F \min_{ij} \|v^{(i)} - v^{(j)}\|^2)$. The best results understood as the maximal degree of expert's acceptance were obtained for the CS coefficient.

4 Generating Fuzzy Rules by Learning from Examples

Many methods have been proposed to generate fuzzy rules by learning from numerical data [5, 2]. In this paper, for generating fuzzy rules by learning from examples, a straightforward algorithm proposed by Wang and Mendel [8] is presented. The task here is to generate a set of fuzzy rules from training data, and use them to determine a mapping $O^{(i)} = f(x_1^{(i)}, x_2^{(i)}, x_3^{(i)})$. The method consists of the following four steps:

1. *Generate a set of membership functions for each input of the feature space*
Generating the set of membership functions begins with variances computation $\text{var}^{(j)}$, $j = 1, 2, \dots, C$ for all clusters resulting from Gustaffson-Kessel clustering presented above. Let $\text{var}^{(j)} = [\text{var}_1^{(j)}, \text{var}_2^{(j)}, \text{var}_3^{(j)}]^T$, then we can make use of the cluster prototypes $v^{(j)}$ and $\text{var}^{(j)}$, which reflect the actual data distribution in the input space, to generate membership functions for each input. The idea is to approximate each cluster as a hyper-ellipsoid with

its center being cluster prototype and the length of axes decided by the corresponding variance of the cluster. The projection of the hyper-ellipsoid onto each axis will produce a symmetric triangular membership function with the peak point being the corresponding component of the cluster prototype $v^{(j)}$ and its variances $\text{var}^{(j)}$. In our case we have described the triangular function as the special case of the trapezoidal function described as:

$$\Pi(x, a, b, c, d) = \begin{cases} 0 & x \leq a \\ (x - a)/(b - a) & a < x \leq b \\ 1 & b < x \leq c \\ (d - x)/(d - c) & c < x \leq d \\ 0 & x > d \end{cases} \quad (7)$$

where $a_q = v_q^{(j)} - h_q^{(j)}/2$, $b = c = v_q^{(j)}$, $d = v_q^{(j)} + h_q^{(j)}/2$ for $q = 1, 2, 3$, $j = 1, 2, \dots, C$ and $h_q^{(j)} = 4 \cdot \text{var}_q^{(j)}$. After generation of the set of membership functions the process of merging neighboring membership functions is realized. If two neighboring membership functions are $\Pi(x, a_{l-1}, b_{l-1}, c_{l-1}, d_{l-1})$ and $\Pi(x, a_l, b_l, c_l, d_l)$, they will be merged if the following condition is satisfied:

$$0.5 \cdot |b_l + c_l - b_{l-1} - c_{l-1}| \leq l_T, \quad (8)$$

where $l_T = [10, 0.1, 0.1]^T$ are a pre-specified thresholds defined for each input. The new membership function after the combination is the following:

$$\Pi(x, \min(a_l, a_{l-1}), \min(b_l, b_{l-1}), \max(c_l, c_{l-1}), \max(d_l, d_{l-1})). \quad (9)$$

As the result of the merging process, some membership functions have trapezoidal shapes instead of triangular ones.

2. *Find the intervals for each input of the feature space*

Find the domain intervals for each input by searching the crossing points between the membership functions generated in the previous step. It makes partition of each input into N_q regions where $q = 1, 2, 3$, denoted as R_1, R_2, \dots, R_{N_q} , and generates a crisp set of cubes in the feature space.

3. *Generate fuzzy rules from given data pairs*

Produce a rule from each input-output data pair included in the training data, for example: $x_1^{(1)} = 120$, $x_2^{(1)} = 0.8$, $x_3^{(1)} = 0.6$; $O^{(1)} = 1 \Rightarrow x_1^{(1)}(u_1^{(1)} = 0.65 \text{ in } R_2(\text{max})), x_2^{(1)}(u_2^{(1)} = 0.75 \text{ in } R_4(\text{max})), x_3^{(1)}(u_3^{(1)} = 0.91 \text{ in } R_3(\text{max})); O^{(1)} = 1 \Rightarrow$

Rule1 : IF $x_1^{(i)}$ is R_2 AND $x_2^{(i)}$ is R_4 AND $x_3^{(i)}$ is R_3 THEN O is Class = 1.

4. *Minimization of fuzzy rules*

As mentioned above, each data pair generates one rule. Usually there are a large number of data pairs (about few thousands) is available, so it is very likely that some conflicting rules are produced. The conflicting rules have the same IF part but different THEN parts. On way to solve this problem one need to assign a soudness degree and from the subset of conflicting rules with

the same IF part accept only one rule with the maximal soudness degree. The soudness degree $SD(Rule_k)$ for k -th rule is defined as:

$$SD(Rule_k) = N_{Rule}/N_{IF_k}, \quad (10)$$

where N_{Rule} denotes the number of data pairs in the training data which support this rule, and N_{IF_k} is the total number of patterns which have the same IF part. By using this frequency based degree, we incorporate the statistical information into fuzzy system resulting in more reliable decision making. After the minimization of fuzzy rules we create rule bank which later will be used for classifying. The proposed system learning for the mean number of objects equals to 1000–4000 which in described above feature space finds 2-3 clusters (classes) generates rule bank including 10–20 rules.

5 Mapping between Input and the Output by Using a Defuzzification Method

The fuzzy rule-based system built in the previous stage can start to realize its task i.e. object classifying. To determine the mapping $O^{(i)} = f(x_1^{(i)}, x_2^{(i)}, x_3^{(i)})$ in fuzzy systems, where $O^{(i)}$ denotes crisp value i.e. class number, a defuzzification method should be adopted. In this system we propose using of the following centroid defuzzification formula [5, 3] to determine the output for each input pattern:

$$O^{(i)} = \frac{\sum_{k=1}^K IF_k^{(i)} O_k}{\sum_{k=1}^K IF_k^{(i)}}, \quad (11)$$

where K is the number of rules, O_k is the class number generated by rule k (O_k takes values of $0, 1, \dots, C-1$ and $IF_k^{(i)}$ is defined as: $IF_k^{(i)} = \prod_{q=1}^3 u_{qk}$ and u_{qk} denotes the membership grade of q -th feature in the fuzzy regions that k -th rule occupies.

6 Experimental Results and Future Work

We have applied the proposed fuzzy rule-based system for classifying nuclei on the contact endoscopy images. The system was tested on an image database of 255 grey scale images. A computer program in the first stage segments input image and producing binary image with objects painted by white. The system have has been taught and tested using five different subsets of the training data where every subset was built with five images including total number of objects equal to 1000–4000. Every subset has included normal and abnormal objects. Three features as described in second section were extracted for each object. Experimental results on the database of 25 images showed that the technique achieved good and reliable results. In our experiments we had learnt the system with every subset and later have tested classification using the remaining four subsets. Comparing classification results, our technique gives similar results

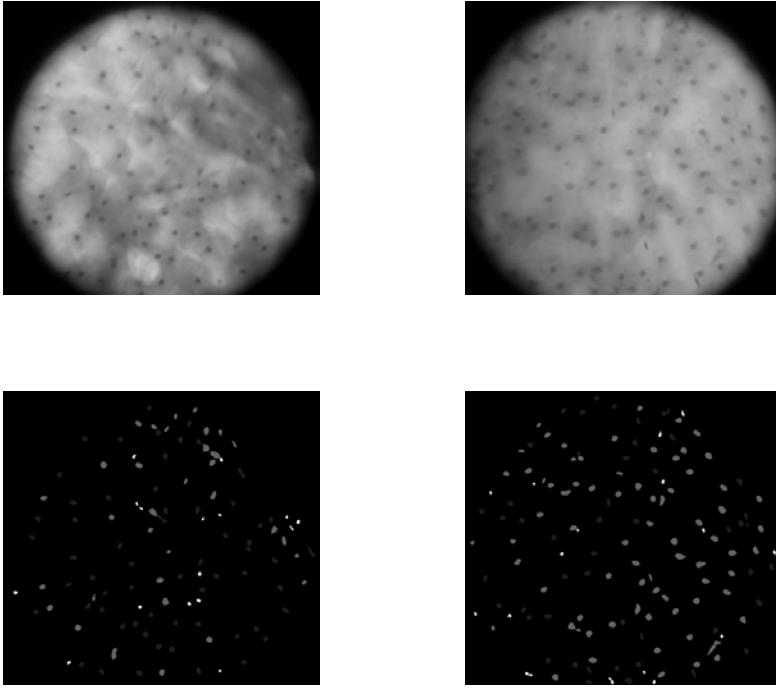


Fig. 1. Contact endoscopy images (top) and their classification results (bottom)

regardless of subset used for training. However, no quantitative studies of its diagnostic accuracy not yet exist, and in the in the near future the classification results will be confirmed in comparison with histopathology images defined in our research as the "golden standard".

Figure 1 shows two original contact endoscopy images with a size 1072×804 and their classification result with three classes. The middle-gray objects were classified to the class called abnormal, dark and white objects define classes normal. Note that the number of white objects belonging is very small and these objects in the clustering terminology would represent the special case of objects called outliers.

Acknowledgement. This work was financially supported by the Polish national budget for scientific research in 2005–2008 within a grant No 3 T11E 01128, what is gratefully acknowledged.

References

1. Andrea, M., Dias, O., Santos, A.: Contact endoscopy during microlaryngeal surgery: a new technique for endoscopic examination of the larynx. *Ann Otol Rhinol Laryngol* 104(5), 333–339 (1995)
2. Bezdek, J.C.: Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York (1981)

3. Chi, Z., Yan, H.: Image segmentation using fuzzy rules derived from K -means clusters. *Journal of Electronic Imaging* 4(2), 199–206 (1995)
4. Gustaffson, W.C., Kessel, W.C.: Fuzzy clustering with a fuzzy covariance matrix, pp. 761–766. IEEE CDC, San Diego, California (1979)
5. Hoppner, F., Klawonn, F., Kruse, R., Runkler, T.: Fuzzy cluster analysis, methods for classification, data analysis and image recognition. John Wiley and Sons, Chichester (1999)
6. Jain, K.A., Dubes, R.C.: Algorithms for clustering data. Prentice Hall, New Jersey (1988)
7. Tarnawski, W., Kurzynski, M.: An improved diffusion driven watershed algorithm for image segmentation of cells. In: Proceedings of the XII International Conference Medical Informatics & Technologies (2007)
8. Wang, L., Mendel, J.M.: Generating fuzzy rules by learning from examples. In: Proceedings of the IEEE International Symposium on Intelligent Control, San Francisco, USA, pp. 38–43 (1991)
9. Wardrop, P.J., Sim, S., McLaren, K.: Contact endoscopy of the larynx: a quantitative study. *J. Laryngol Otol.* 114(6), 437–440 (2000)

Synthesis of Medical Images in the Domain of Melanocytic Skin Lesions

Zdzisław S. Hippe¹, Jerzy W. Grzymała-Busse^{1,2}, and L. Piątek³

¹ Department of Expert Systems and Artificial Intelligence, University of Information Technology and Management, 35-225 Rzeszów, Poland

zhippe@wsiz.rzeszow.pl

² Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS 66045, US

Jerzy@ku.edu

³ Department of Distributed Systems, University of Information Technology and Management, 35-225 Rzeszów, Poland

lpiatek@wsiz.rzeszow.pl

Summary. In this paper, the development of a new module of Internet Information System for synthesis of melanocytic skin lesion images is briefly outlined. The key approach in the developed synthesis methodology of images is a semantic conversion of textual description of melanocytic skin lesions - by an in-house developed system - into digital images. It was found, that the developed methodology can be successfully used in the process of teaching of dermatology students and also in training of preferred medical doctors. The system is available via Internet at the website <http://www.melanoma.pl>.

1 Introduction

Increasing subspecialization in medicine results in an easy availability of medical doctors at medical centers located in large cities, and a lack them in small, predominantly rural areas. Physicians in these areas may have problems with diagnosing of some pigmented skin lesions. For this reason, in many cases, patients are moved to medical centers in neighbour city for adequate treatment. This transportation means inconvenience for patients, and unexpected costs for the health care system.

In the last few years continuous progress in information technology has led to broader use a revolutionary diagnostic procedure, based on the concept of telemedicine. This concept improves communication between physicians and medical specialists, and helps in reducing costs of medical treatments. One of important parts of current telemedicine is teledermatology (also known as teledermoscopy [1]). It enables the sending of dermoscopic images of pigmented skin lesions over telemetric networks [2], and a fast comparison of diagnosis of the same skin lesions by different experts. Another encouraging approach due to the fast and easy exchange of information *via* the Internet is the possibility of discussing dermoscopic images with experts from many countries. An impressive example would be the New York University Group's "e-Room" project, called

Dermnetwork (<http://www.dermnet-work.org>) where information on interesting dermoscopy cases can be shared.

In our research we focus the attention on the development of effective algorithms for semantic conversion of textual description of melanocytic lesions into respective images. In other words, the main goal of our research constitutes in elaboration also called generator of images. We follow this concept, because according to the personal data protection act, both making and publishing of real photographs of melanocytic lesions requires patient’s approval. This could obstruct the creation of informational databases, used by less experienced medical doctors. It was assumed, that by the application of the generator of images, can intent on the constraint of the use of real digital pictures in favour of synthesized images, representing symptoms of melanocytic lesions with required precision.

2 Structure of the Source Dataset

In our research we use the source informational database, initially described in [3]. Data contained in this database were collected in the Outpatient Center of Dermatology in Rzeszow, Poland. Each case is described by **13** attributes, divided into **4** groups. These attributes were: *asymmetry*, character of the *border* of a lesion, combination of *colors* and combination of *structures* observed in the lesion. To finish with, an additional attribute was used, namely the **TDS**-parameter (**T**otal **D**ermatoscopy **S**core) [4]. It was computed for each case (patient), using the formula described later. Each investigated case was assigned to one of four possible decision categories: *benign nevus*, *blue nevus*, *suspicious nevus* or *melanoma malignant*. The number of different diagnoses in the database are shown in Table 1.

Table 1. The number of different diagnoses in the database

Diagnose	Number of cases	%
<i>Benign nevus</i>	248	45
<i>Blue nevus</i>	78	14
<i>Suspicious nevus</i>	108	20
<i>Melanoma malignant</i>	114	21

In our database the attribute *Asymmetry* defines the symmetry of a lesion along two axes, crossed at a right angle [5]. Logical values of this attribute can be: *symmetric change* (numeric value in the base = **0**), *one-axial asymmetry* (numeric value = **1**) and *two-axial asymmetry* (numeric value = **2**). Description the character of the rim of a lesion is based on splitting it into eight equal parts by four axes crossed in a point, and assigning **0** or **1**, if the border between a lesion and the skin is **diffuse** or **sharp**, respectively. Therefore, the value of the *Border* attribute oscillates between **0** and **8**. Then, the attribute *Color* can have **6** allowed values: *black*, *blue*, *dark-brown*, *light-brown*, *red* and *white*. On the

other hand, the attribute *Diversity* can have 5 logical values: *branched streaks*, *pigment dots*, *pigment globules*, *pigment network* and *structureless areas*. The **TDS** value is computed according to the **ABCD rule**:

$$TDS = 1,3 * \mathbf{A}symmetry + 0,1 * \mathbf{B}order + 0,5 * \sum \mathbf{C}olor + 0,5 * \sum \mathbf{S}tructure \quad (1)$$

where **A** means a description of lesion's asymmetry, **B** is a description of lesion's border, **C** is a specification of combinations of colors appearing in considered lesion, and **D** is a specification of lesion's diversity. In this way, verification of our research was possible with use of the concept of constructive induction [6].

Recently, synthesis of discussed images is based on latest results of the research executed at Kansas University [7, 8], differentiating the role of a particular color and structure within a lesion, allowing to determine the value of a new total dermatoscopy score parameter, called **New_TDS**:

$$\begin{aligned} TDS = & (0,8 * Asymmetry) + (0,11 * Border) + (0,5 * C_White) \\ & + (0,8 * C_Blue) + (0,5 * C_DarkBrown) \\ & + (0,6 * C_LightBrown) + (0,5 * C_Black) \\ & + (0,5 * C_Red) + (0,5 * Pigment_Network) \\ & + (0,5 * Pigment_Dots) + (0,6 * Pigment_Globules) \\ & + (0,6 * Branched_Streaks) + (0,6 * Structureless_Areas) \end{aligned} \quad (2)$$

3 Methodology of the Research

Developed methodology of generation of images consists of the composition of predefined fragments of images of melanocytic lesions (so called textures). Such procedure seems to be quite satisfactory in relation to two characteristic attributes of images i.e. *Color* and *Diversity* of lesions. The simulation of *Asymmetry* of the lesions and character of their rim (*Border* attribute) requires the special approach, based on random selection of allowed logical values for these attributes and combining them in an exhaustive way into a set of simulated images. According to the present concept, each vector from the source (textual) database, describing an anonymous patient, can produce a collection of simulated images. These images, according to Kulikowski [9], could be treated as synonyms in the space of images.

4 Synthesis of Lesion's Asymmetry

The problem of asymmetry is solved by putting a combination of four basic structures, created from a priori prepared parts (of the size 200*200 pixels) (see Fig. 1).

All attributes of a texture that stands for a constructing part (a quarter) of an image, are different due to the diversity of lesion asymmetry. It should be

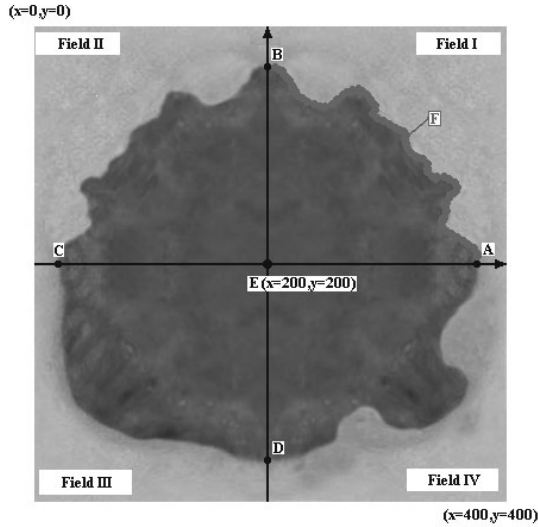


Fig. 1. Partition of an image and arrangement of its parts

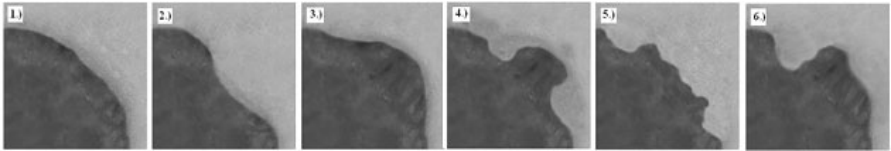


Fig. 2. Pre-defined fragments of images (quarters), with a different form of the rim (curve **F**)

mentioned that the four components of a simulated image differ each other only by a shape of the curve **F** (see Fig. 1). Due to various shapes of this curve **F** there are different types of construction elements, responsible for the simulation of asymmetry (see Fig. 2).

There are 3 possibilities of mapping the *Asymmetry* attribute:

- in the first one, for the *symmetric change* there is only one way of mapping a fragment randomly selected of some possible fragments (various shapes of the **curved F**), which is next placed in each of four fields of the main square,
- for the *one-axial asymmetry change*, two different textures are randomly chosen from six a priori defined ones. Next, one of them is placed in the fields no. **I** and **IV**, whereas the other one is placed in the fields **II** and **III**,
- finally, for the *two-axial asymmetry change*, three different construction parts (textures) are chosen, and then one of them (randomly selected) is repeated onto the fields labeled **I** and **II**, whereas the other two fragments are placed (randomly too) in the remaining free fields **III** and **IV**.

After placing all four construct fragments in particular fields of the main square (Fig. 1), it's received an image characteristic for the type of asymmetry, described by a given textual data vector from the source dataset.

5 Synthesis of Lesion's Border

Algorithm of synthesis of lesion's **Border** is similar as diagnosing this symptom by medical doctors, relies on splitting of a lesion into 8 regular parts, and then counting how many of them displayed **sharp** transition towards the skin (count=1), and how many displayed **diffuse** transition (count=0), so the numerical value of this symptom is in the range from 0 to 8. In this way distribution of possible "isomers" of transforms increasing to 256 possibilities of sharp/diffuse transitions. Each combination of transitions (except for *border* equal 0, 1 or 8) is multiplied by 8, because a set of 8 new transitions can be generated applying the operation of eight-fold symmetry axis, perpendicular to the plane of the lesions image. Transitions for the *border*=1 and *border*=8 can be treated as distinct representations, applying the approach of Schoenflies points groups [10]. All of them are then applied in the superposition with previously simulated asymmetry of melanocytic lesions.

6 Synthesis Colors and Structures of the Lesion's

Synthesis of colors and structures of lesion's should consider multi-argument character of *Color* and *Diveristy of structure* attributes, capable to create considerable number of combinations of these parameters, which can simultaneously appear in a given lesion. *Color* can have 6 allowed values: *black, blue, dark-brown, light-brown, red* and *white*, when at the same time the *Diversity of structure* attribute can have 5 logical values: *branched streaks, pigment dots, pigment globules, pigment network* and *structureless area*. Calculations of all

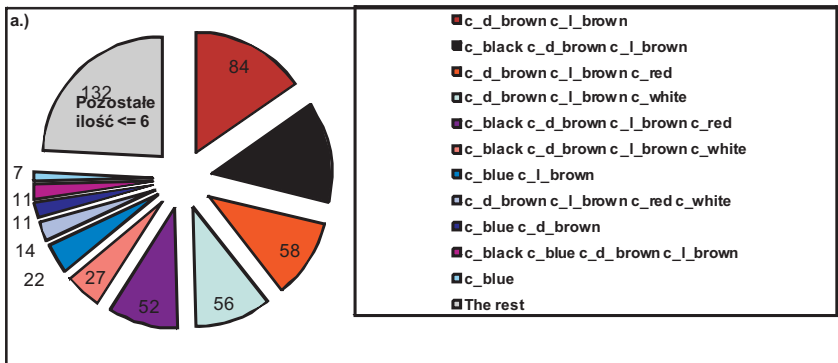


Fig. 3. Simultaneous occurrence of colors and structures in real lesion's images

possible combinations of those both features required a special approach: initially (before generating textures), there is an attempt to find which colors and structures occur simultaneously in a real lesions [11]. The results of these investigations are shown in Fig. 3.

Basing on these findings, the developed algorithms can be simplified by application of so called basic textures, having two combinations of *colors* only: images with color blue, appearing separately, and images with colors dark-brown and light-brown. The remaining colors in combination with required diversity of structures are added by the dynamic application of so called remaining layers.

7 Program Implementation

The discussed system, called by us a generator of images of melanocytic lesions (see Fig. 4), is implemented with using the language **PHP** (Programming Hypertext Preprocessor), combined with use of graphic library **GD** [12]. Pre-defined textures, necessary for the reasonable simulation (202 in number) of lesion, were defined in **PNG** (Portable Network Graphics) format, where each texture contains various number of combinations of descriptive features.

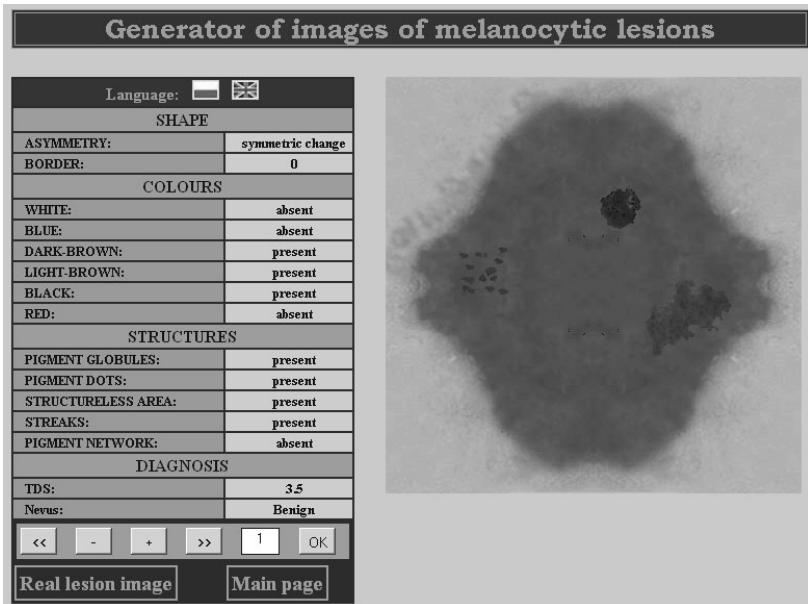


Fig. 4. Generator of images of melanocytic lesions *via* Internet

8 Summary and Conclusion

A new module of internet-based information system for classification of melanocytic skin lesions, is briefly here described. The developed algorithms

enables to generate the exhaustive number of synthesized images, corresponding to symptoms contained in a given lesion, originally described by textual vector from the source database. Synthesized images enable to create multi-category informational database, which can be successfully used not only in teaching of medicine students, but also in job practice of dermatologists and preferred medical doctors. This system is available on the Internet website: <http://www.melanoma.pl>.

References

1. Piccolo, D., Smolle, J., Wolf, I., Peris, H., Hofmann-Wellenhof, R., Dell'Eva, G., Burrioni, M., Chimenti, S., Kerl, H., Soyer, H.: "Face-to-face" versus remote diagnosis of pigmented skin tumors: a teledermoscopic study. *Arch Dermatol* 135, 1467–1471 (1999)
2. Provost, N., Kopf, A., Rabinovitz, H., Stolz, W., De David, M., Wasti, Q., Bart, R.: Comparison of conventional photographs and telephonically transmitted compressed digitized images of melanomas and dysplastic nevi. *Dermatology* 196, 299–304 (1998)
3. Hippe, Z.S.: Computer Database 'NEVI' on Endangerment by Melanoma. *TASK Quarterly* 3(4), 483–488 (1999)
4. Hippe, Z.S., Bajcar, S., Bśajdo, P., Grzymała-Busse, J.P., Grzymała-Busse, J.W., Knap, M., Paja, W., Wrzesień, M.: Diagnosing Skin Melanoma: Current versus Future Directions. *TASK Quarterly* 7(2), 289–293 (2003)
5. Bajcar, S., Grzegorzczak, L.: *The Atlas of Diagnostics of Melanocytic Lesions*. Jagiellonian University Editorial Office, Cracow (Poland) (2002)
6. Michalski, R.S., Bratko, I., Kubat, M. (eds.): *Machine Learning and Data Mining: Methods and Applications*. J. Wiley & Sons Ltd, Chichester (1998)
7. Alvarez, A., Bajcar, S., Brown, F.M., Grzymała-Busse, J.W., Hippe, Z.S.: Optimization of the ABCD Formula Used for Melanoma Diagnosis. In: Kłopotek, M.A., Wierzchoń, S.T., Trojanowski, K. (eds.) *Advances in Soft Computing (Intelligent Information Processing and Web Mining)*, pp. 233–240. Springer, Berlin (2003)
8. Grzymała-Busse, J.W., Hippe, Z.S.: Data Mining Methods Supporting Diagnosis of Melanoma. In: *Proc. of the 18th IEEE Symposium on Computer-Based Medical Systems*, pp. 371–373. IEEE Comp. Soc, Los Alamitos (2005)
9. Kulikowski, J.L.: The foundations of the structural description of distracted databases of expert knowledge. In: *Proc. of the Conference Databases for Science and Technology, Gdańsk (Poland), September 25-27, 2005*, pp. 29–38 (in Polish, 2005)
10. Alberty, R.A., Silbey, R.J.: *Physical Chemistry*, pp. 426–430. J. Wiley & Sons Ltd, New York (1992)
11. Hippe, Z.S., Piątek, L.: From research on the database of simulated medical images. In: *Proc. of the Conference Databases for Science and Technology, Gdańsk (Poland), September 25-27, 2005*, pp. 225–230 (2005)
12. <http://www.boutell.com/gd/> (January 15, 2008)

Identification of Layers in a Tomographic Image of an Eye Based on the Canny Edge Detection

Robert Koprowski and Zygmunt Wrobel

University of Silesia, Faculty of Computer Science and Materials Science Institute of Computer Science, Department of Biomedical Computer Systems. ul. Bedzinska 39, 41-200 Sosnowiec
koprow@us.edu.pl, wrobel@us.edu.pl

Abstract. In the paper we present an algorithm for the identification of retina layers using the Canny edge detection for images obtained with OCT (Optical Coherence Tomography) Copernicus. The developed algorithm is an extension of the approaches covered in [4] and allows the identification and detection of hyaline-retinal border layers, the retina and other. The algorithm was implemented in the Matlab environment and the C language.

1 Introduction

The L_{GRAY} input image is prefiltered with a median filter with mask ($M_h \times N_h$) of size $h = 3 \times 3$. The obtained image L_{MED} is filtered again with a modified Canny filter; the consecutive stages of this process are presented in the following chapters.

2 Canny Filtering

The first stage of the edge detection method [1, 3, 6] is to perform the convolution on the input image L_{MED} , i.e.:

$$L_{GX}(m, n) = \sum_{m_h=-M_h/2}^{M_h/2} \sum_{n_h=-N_h/2}^{N_h/2} L_{GRAY}(m + m_h, n + n_h) \cdot h_x(m_h, n_h) \quad (1)$$

$$L_{GY}(m, n) = \sum_{m_h=-M_h/2}^{M_h/2} \sum_{n_h=-N_h/2}^{N_h/2} L_{GRAY}(m + m_h, n + n_h) \cdot h_y(m_h, n_h) \quad (2)$$

with the following Gaussian filter masks 3×3 :

$$\begin{bmatrix} 0.325 & 0.536 & 0.325 \\ 0 & 0 & 0 \\ -0.325 & -0.536 & -0.325 \end{bmatrix} \quad \begin{bmatrix} 0.325 & 0 & -0.325 \\ 0.536 & 0 & -0.536 \\ 0.325 & 0 & -0.325 \end{bmatrix}$$

The gradient matrix (for both directions) required for determining edges was chosen using the typical formula:

$$L_{GXY}(m, n) = \sqrt{L_{GX}(m, n)^2 + L_{GY}(m, n)^2} \quad (3)$$

And threshold p_{xy} :

$$p_{xy} = \varepsilon \cdot \left(\max_{m,n \in LG_{XY}} (L_{G_{XY}}(m,n)) - \min_{m,n \in LG_{XY}} (L_{G_{XY}}(m,n)) \right) + \min_{m,n \in LG_{XY}} (L_{G_{XY}}(m,n)) \quad (4)$$

where ε is a coefficient selected in the $(0, 1)$ range. In order to obtain the final form of the formula for the matrix of the image containing the edges (L_{BIN_KR}), another steps are necessary - defining $L_{G_{XYM}}$, i.e.:

$$L_{G_{XYM}}(m,n) = \begin{cases} p_{xy} & \text{if } L_{G_{XY}}(m,n) < p_{xy} \\ L_{G_{XY}}(m,n) & \text{if } L_{G_{XY}}(m,n) \geq p_{xy} \end{cases} \quad (5)$$

and obtaining the coordinates of i_{xy} and j_{xy} , (x_i, y_i) and (x_j, y_j) , respectively, determined by using the formula:

$$x_i = \cos(\alpha(m,n)) \text{ and } x_j = -\cos(\alpha(m,n)) \quad (6)$$

$$y_i = \sin(\alpha(m,n)) \text{ and } y_j = -\sin(\alpha(m,n)) \quad (7)$$

where the α angle was calculated for each L_{GX} and L_{GY} pixel pair:

$$\alpha(m,n) = \text{atan} \left(\frac{L_{GY}(m,n)}{L_{GX}(m,n)} \right) \quad (8)$$

and eventually obtaining the i_{xy} and j_{xy} values, which assume the saturation level according to the values interpolated on the surface determined from the area of the 3×3 resolution with $L_{G_{XYM}}(m \pm \Delta m, n \pm \Delta n)$, where Δm and Δn are equal to 1 (see fig.1 and 2).

Consequently, the L_{BIN_KR} output image of the edges determined with the Canny method is:

$$L_{BIN_KR}(m,n) = \begin{cases} 0 & \text{then } L_{G_{XYM}}(m,n) \leq p_{xy} \\ 1 & \text{then } (L_{G_{XYM}}(m,n) > p_{xy}) \wedge (L_{G_{XYM}}(m,n) > i_{xy}) \wedge \\ & \wedge (L_{G_{XYM}}(m,n) > j_{xy}) \\ 0 & \text{then } (L_{G_{XYM}}(m,n) > p_{xy}) \wedge ((L_{G_{XYM}}(m,n) \leq j_{xy}) \vee \\ & \vee (L_{G_{XYM}}(m,n) \leq i_{xy})) \end{cases} \quad (9)$$

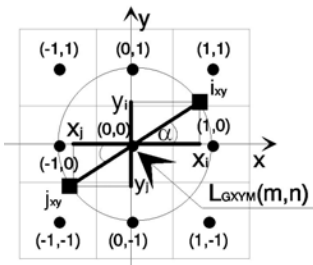


Fig. 1. Graphical interpretation of the i_{xy} and j_{xy} point location in the $L_{G_{XYM}}(m \pm 1, n \pm 1)$ image fragment

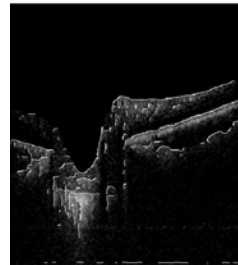


Fig. 2. L_{MED} input image and the white pixels of the L_{BIN_KR} image

The example image OCT shown in fig. 2 was obtained for $e=0.15$; for the sake of the better result evaluation the white pixels of the L_{BIN_KR} image were shown.

3 Edge Line Properties

For the L_{BIN_KR} image a labeling operation was performed during which each cluster (of "1" values) was given a label $e_t = 1, 2, \dots, Et - 1, Et$. Then, for every label e_t , a dilation operation is performed with a rectangular structural element SE_d of the 5×1 size oriented according to the value of the $\alpha(m, n)$ angle, where the origin of the coordinate system is located in the first row of the element. The L_{IND} result image in pseudocolor is presented in fig. 3.

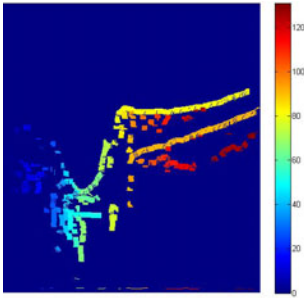


Fig. 3. L_{IND} image in pseudocolor (the number of labels $Et = 131$)

e_t	P_{e_t}	I_{e_t}
1	509	0.28
2	42	0.29
3	88	0.15
4	20	0.11
5	13	0.17
6	16	0.14
7	40	0.08
8	74	0.11
9	35	0.34

Fig. 4. Table of weights with examples of values for the objects with first several e_t labels

In fig. 4, we present the weight values for several consecutive first labels from the L_{IND} image (fig. 3), i.e. the L_{e_t} binary images where P_{e_t} is the area of the object for label e_t and I_{e_t} is the average value of its grayscale level:

$$P_{e_t} = \sum_{m=1}^M \sum_{n=1}^N L_{e_t}(m, n) \tag{10}$$

$$I_{e_t} = \frac{1}{M \cdot N} \cdot \sum_{m=1}^M \sum_{n=1}^N (L_{e_t}(m, n) \cdot L_{MED}(m, n)) \tag{11}$$

The determined values P_{e_t} and I_{e_t} are employed as properties during the final analysis of edge lines.

4 Modified Active Contour

Every continuous edge line visible in the L_{e_t} image for labels $e_t = 1, 2, \dots, Et - 1, Et$ is transformed into a vector x_{e_t} and yet that represents the coordinates of

points in the Cartesian coordinate system. The modified active contour method is applied in order to "elongate" each of the edges in both directions [2, 3, 7]. With this aim in view, for the first two coordinate pairs of the first edge $(x_1(1), y_1(1))$ and $(x_1(2), y_1(2))$ and for the last two ones $(x_1(end - 1), y_1(end - 1))$ and $(x_1(end), y_1(end))$ a straight line that passes through these points (end stands for the end element) is determined, according to the figure below (fig. 5):

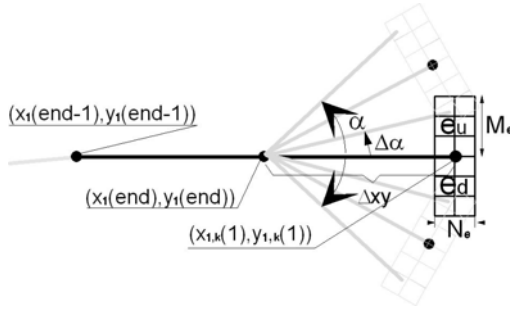


Fig. 5. The graphical interpretation of the modified active contour method used for determining the consecutive points, starting with the location of the points $(x_1(end - 1), y_1(end - 1))$ and $(x_1(end), y_1(end))$ for the new selected point (pixel) $(x_{1,k}(1), y_{1,k}(1))$. For simplification, the inclination angle of the end points of the edge was set to $\beta = 0^\circ$.

In fig. 5, we present the principle of the active contour method in which, starting from points $(x_1(end - 1), y_1(end - 1))$ and $(x_1(end), y_1(end))$ a straight line with inclination angle β_1 that passes them is determined, i.e.:

$$\beta_1(x_1(end), y_1(end)) = atan \left(\frac{y_1(end) - y_1(end - 1)}{x_1(end) - x_1(end - 1)} \right) \quad (12)$$

and, at the Δxy distance the location of a new point $(x_{1,k}(1), y_{1,k}(1))$ is determined for various possible locations (within angle range $\beta_1(1) \pm \alpha$ every $\Delta \alpha$). The selection of the correct location of the point of the contour (created by adding consecutive points to the existing edge) is achieved based on the analysis of the average value from the $e_{u1}(x_u, y_u, \alpha, 1)$ and $e_{d1}(x_d, y_d, \alpha, 1)$ areas of the $M_e \times N_e$ size. For each location of point $(x_{1,k}(1), y_{1,k}(1))$ difference ΔS is calculated:

$$\Delta S(1, \alpha) = \frac{1}{M_e \cdot N_e} \cdot \left(\sum_{y_u=1}^{M_e} \sum_{x_u=1}^{N_e} e_u(x_u, y_u, \alpha, 1) \cdot h_u(x_u, y_u) - \sum_{y_d=1}^{M_e} \sum_{x_d=1}^{N_e} e_d(x_d, y_d, \alpha, 1) \cdot h_d(x_d, y_d) \right) \quad (13)$$

Where: x_u, y_u - the coordinates of the consecutive elements of the e_u i h_u matrix located above the analysed point $(x_{1,k}(1), y_{1,k}(1))$ for which $x_u \in \{1, 2, \dots, N_u - 1, N_u\}$ and $y_u \in \{1, 2, \dots, M_u - 1, M_u\}$.

x_d, y_d - the coordinates of the consecutive elements of the e_d i h_d matrix located below the analysed point $(x_{1,k}(1), y_{1,k}(1))$ for which $x_d \in \{1, 2, \dots, N_d - 1, N_d\}$ and $y_d \in \{1, 2, \dots, M_d - 1, M_d\}$ The masks h_u i h_d for $M_e \times N_e = 3 \times 2$:

$$h_u = \begin{bmatrix} 0.5 & 0.5 \\ 0.7 & 0.7 \\ 1 & 1 \end{bmatrix} \qquad h_d = \begin{bmatrix} 1 & 1 \\ 0.7 & 0.7 \\ 0.5 & 0.5 \end{bmatrix}$$

The areas (matrices) e_u and e_d of the $M_e \times N_e$ size are created based on the β and α angle every $\Delta\alpha$ in the following way:

$$e_{u1}(x_u, y_u, \alpha, 1) = L_{MED}(y_{1,k}(1) + y_u \cdot \cos(\beta_1(1) + \alpha + 90), \qquad (14)$$

$$\qquad \qquad \qquad , x_{1,k}(1) + x_u \cdot \sin(\beta_1(1) + \alpha + 90))$$

$$e_{d1}(x_d, y_d, \alpha, 1) = L_{MED}(y_{1,k}(1) + y_d \cdot \cos(\beta_1(1) + \alpha + 90), \qquad (15)$$

$$\qquad \qquad \qquad , x_{1,k}(1) + x_d \cdot \sin(\beta_1(1) + \alpha + 90))$$

where: $x_u \in \{1, 2, 3, \dots, N_e - 1, N_e\}$ and $y_u \in \{1, 2, 3, \dots, M_e - 1, M_e\}$ and $\beta_1(1)$, in general $\beta_1(v_1)$:

$$\beta_1(v_1) = \text{atan} \left(\frac{y_{1,k}(v_1) - y_{1,k}(v_1 - 1)}{x_{1,k}(v_1) - x_{1,k}(v_1 - 1)} \right) \qquad (16)$$

for $v_1 \in \{2, 3, \dots, V_1 - 1, V_1\}$, V_1 - the total number of points in the contour line. The angle for which there is a best fit for the analysed point $(x_{1,k}(v_1), y_{1,k}(v_1))$ is calculated as α^* for which $\Delta S(v_1, \alpha)$ reaches its maximum or minimum depending on the location and brightness of the analysed object.

$$\Delta S(v_1, \alpha^*) = \max_{\alpha}(\Delta S(v_1, \alpha)) \qquad (17)$$

The consecutively determined points for increasing v_1 need to be limited. The condition is the minimal value of $\Delta S(v_1, \alpha^*)$ with the pr threshold.

The proposed modified active contour method has very interesting properties. The parameters of this part of the algorithm include:

- α - the angle specifying the range in which the best fit according to the selected criterion is searched for,
- $\Delta\alpha$ - the resolution with which the best fit is searched for,
- Δxy - the distance between the current and next active contour point (i.e. one that is searched for),
- M_e - the height of the analysed e_u and e_d areas,
- N_e - width of the analysed e_u and e_d areas.

Below (fig. 6- fig.9) we presented the results for a test image of a square for the aforementioned parameters $\alpha, \Delta\alpha, \Delta xy, M_e, N_e$ modified in the range $\alpha \in \{1, 2, 3, \dots, 19, 20\}$, $\Delta xy = N_e \in \{1, 2, 3, \dots, 19, 20\}$, $M_e \in \{1, 2, 3, \dots, 19, 20\}$ for $\Delta\alpha = 1$ and $p_r = -0.001$. The number of iterations was limited to 50.

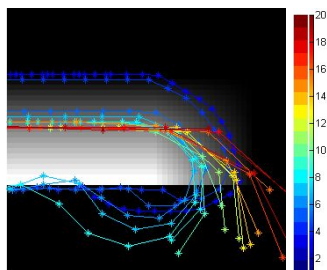


Fig. 6. The test image and a fragment of the results from the modified active contour method for $\alpha = 40$, $\Delta\alpha = 1$, $M_e = 10$ and $\Delta xy = N_e$ modified in the (1, 20) range

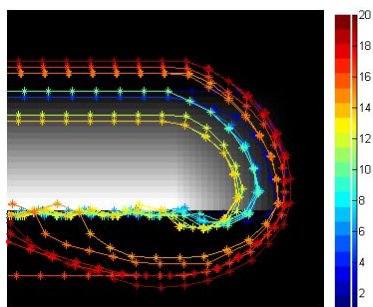


Fig. 7. The test image and a fragment of the results from the modified active contour method for $\alpha = 40$, $\Delta\alpha = 1$, $\Delta xy = N_e = 4$ and M_e modified in the (1, 20) range

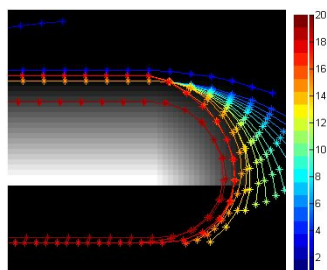


Fig. 8. The test image and a fragment of the results from the modified active contour method for $\alpha = 40$, $M_e = 10$, $\Delta xy = N_e = 10$ and $\Delta\alpha$ modified in the (1, 20) range

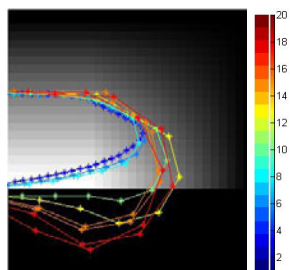


Fig. 9. The test image and a fragment of the results from the modified active contour method for 45, $M_e = 10$, $N_e = 10$, $\Delta\alpha = 1$ and Δxy modified in the (1, 20) range

There are following properties of the presented modified active contour method:

- α - the angle specifies the search range in the sense of the object edge curvature level,
- $\Delta\alpha$ - the resolution with which the curvature level is searched for,
- Δxy - the distance between the current and next point that influence the generalization level and the level of the approximation of intermediate values,
- M_e - the height of the analysed area influencing the algorithm capability to find objects of a higher detail level,
- N_e - the width of the analysed area that averages the searched contour along the edge.

5 The Final Analysis of the Contour Line

The obtained individual contour lines e_t and respective values I_{e_t} and P_{e_t} (the average brightness value and the area) were corrected. In particular, the edges for which $I_{e_t} < p_r \cdot \max_{e_t \in \{1,2,3,\dots,E_t\}}(I_{e_t})$ and these for which $P_{e_t} < p_r \cdot \max_{e_t \in \{1,2,3,\dots,E_t\}}(P_{e_t})$ (where p_r was arbitrarily selected at the 0.2 (20%) level) were removed. On the remaining edges e_k (that is ones that were not removed) the correction was applied by using the active contour method at their ends. The values of the active contour parameters were selected at the following levels: $\alpha = 45^\circ$, $\Delta\alpha = 1$, $\Delta xy = 1$, $M_e = 11$, $N_e = 11$. For the individual e_k edges the iterations were stopped when any of the following occurred:

- the maximum number of iterations was reached (set arbitrarily to 1000),
- the condition $\Delta S(v_{e_k}, \alpha^*) < p_s$ (where p_s was set to 0.02) was not met for that point,
- at least two points have the same coordinates - this prevents infinite loop in the algorithm.

Below (fig. 10, fig. 11) we present results obtained for the parameters selected in the aforementioned way.

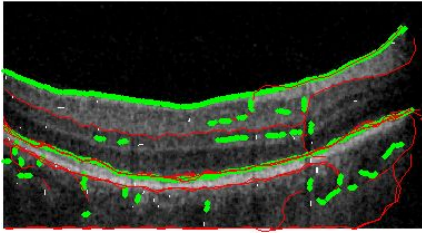


Fig. 10. The effects of the active contour method applied to a real image for $\alpha = 40$, $\Delta\alpha = 1$, $\Delta xy = N_e = 11$, $M_e = 10$. The contour obtained with the Canny method was marked with the green line, the points from the active contour method were marked with the red one

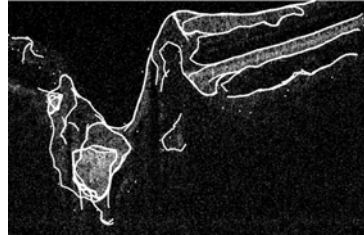


Fig. 11. The effects of the active contour method applied to a real image after the aforementioned correction for $\alpha = 40$, $\Delta\alpha = 1$, $\Delta xy = N_e = 11$, $M_e = 11$

As it can be seen in the figures above (fig. 10, fig. 11), the proposed method correctly determines the individual layers in the OCT image of an eye. The further steps, which we envisage for continuing this approach, are connected with an in-depth analysis of the algorithm from the perspective of the parameter selection.

6 Summary

The presented method that combines the Canny edge detection algorithm and the modified active contour algorithm finds its use in detecting hyaline-retinal

border layers in OCT topographic images of an eye. The proposed method can prove useful for the segmentation of images with different content, provided the appropriate modifications of the algorithm parameters are applied. The obtained results (presented above) are satisfactory. Nonetheless, there is a vast area for the further research on this algorithm, namely optimising it for speed. The next step is the analysis of an image sequence and 3D reconstruction, which will be presented in subsequent papers. Images on fig. 2 and fig. 10 will obtain at Optopol.

References

1. Canny, J.: A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8(6) (November 1986)
2. Davies, E.: *Machine Vision: Theory, Algorithms and Practicalities*, Ch. 5. Academic Press, London (1990)
3. Gonzalez, R., Woods, R.: *Digital Image Processing*, Ch. 4. Addison-Wesley Publishing Company, Reading (1992)
4. Koprowski, R., Wróbel, Z.: Analiza warstw na tomograficznym obrazie oka, *Systemy wspomaganie Decyzji* (2007)
5. Koprowski, R., Wróbel, Z.: Determining the contour of cylindrical biological objects using the directional field. In: *Proc. CORES 2007*, Springer, Heidelberg (2007)
6. Wróbel, Z., Koprowski, R.: *Automatyczne metody analizy orientacji mikrotubul*, Wydawnictwo UŚ (2007)
7. Witkin, A., Terzopoulos, D.: *Int. J. Active Contour Models*, M. Kass. *Computer* 1(4), 321–331 (1987)

Diagnostic Quality-Derived Patient-Oriented Optimization of ECG Interpretation

Piotr Augustyniak

AGH University of Science and Technology, 30 Mickiewicza Ave. 30-059 Krakow
Poland
august@agh.edu.pl

Summary. The paper discusses various aspects of diagnostic quality estimate in an adaptive ECG interpretation system. Since the ECG interpreting software adapts to the patient status and diagnostics goal variations, the commonly used term of data quality must be revised. We propose backgrounds and demonstrate the use of a multidimensional quality hyperspace depending on the assumed system adaptability. The emerging need of appropriate reference database and standardized description of human experts behavior is also addressed here. The subset of proposed quality estimators is used as adaptivity criteria in a prototype of wireless monitoring system for cardiology. The paper concludes with guidelines for testing the processing performance and with a comparison of rigid and adaptive software features.

1 Introduction

Quality control is the aspect of principal importance in every automated support of diagnoses. Since the end-user is not able to fully supervise the behavior of the software and the automatic system should allow him to pay more attention to the patient, professional organizations like cardiology societies implement strict certification procedures for medical electronic equipment [4]. In the domain of automated ECG interpretation software testing, worldwide standard databases with reference signals and data are used to measure whether the values issued by the software under test fall within the tolerance margin around the corresponding reference value (gold standard) [3]. The research towards adaptive distributed ECG interpretation networks revealed unprecedented advantages. Such networks shows high flexibility of interpretation task sharing, eliminating the unnecessary computation and data flow, finally they adapt to the variable status of the patient and diagnostic goals [2]. Until today, there was no standard to test these new features beyond the parameters common to the rigid software. Our proposal aims to fill this gap and to implement a multidimensional hyperspace of quality. Since the target system under test is adaptive and time plays crucial role in life-critical cardiac events, the quality estimation has to support dynamic behavior of the system and include transient description parameters. Unfortunately, medical

guidelines and testing standards (e.g. AHA, IEC) describe only stable pathologies and provide reference records stationary in medical sense. This is sufficient for off-line interpretation systems attempting each part of the signal with the same assumption and thus guaranteeing high repeatability of results. Human experts, however, behave differently taking into account not only a limited section of cardiac electrical image, but also much wider context of history including extra-cardiac events. Staying in touch with their patients, human experts often witness emergencies of pathological events and modify their further diagnostic goal. Design of a remote adaptive interpretation system that is expected to simulate the presence of doctor, must consider new criteria of adaptivity and assessment of diagnostic quality present in the everyday clinical life, but not formally covered by current standards. These criteria should refer to the present patient status and cover:

- specific area of interest for further diagnosis (the optimal hierarchy of diagnostic parameters),
- expected data variability and resulting minimum update frequency of each parameter,
- tolerance of each parameter value,
- possible subsequent diagnoses (patient status) ordered by the likelihood of occurrence,
- reference records containing example transient signals.

2 Materials and Methods

2.1 Adaptive ECG Interpretation System Overview

The developed wireless adaptive ECG monitoring system consists of a star-shaped network managed by a stationary central server receiving medical data and controlling several remote wearable recording devices (fig. 1). The bi-directional GPRS link is used as data carrier, however the long-distance connections use the Internet infrastructure. Wearable recorders are manufactured as low-cost general-purpose instruments for vital signs acquisition. The recorder consists of a battery-operated computer integrated with signal digitizer, communication module and user interface. The hardware design provides wide range of control and re-programmability by the software.

The software consists of a rigid mandatory background and optional overlay modules that can be reconfigured in course of seamless operation. The background contains basic common modules: data acquisition and wireless communication services as well as fundamental user interface procedures. The overlay includes a repository of interpretation and report-formatting procedures programmed as diagnosis-oriented dynamic libraries. The upload and linking or release and deletion of each particular library is performed by the supervising server in any convenient time with respect to other linked libraries and to the available computational power. This approach personalizes the remote recorder to the patient-specific signal features and gives an unprecedented flexibility

required for a pertinent real-time reaction for unexpected events. Most common and frequent cardiac episodes are interpreted by the wearable device software and the result fit in a cost-acceptable data stream. The occurrence of any difficult or unresolved event is reported as a short strip of raw signal for the interpretation by the server software automatically, or even in very rare cases with the assistance of a human expert. The prototype follows human relations-based organization of ECG monitoring particularly in two aspects:

- processing adaptability,
- reporting adaptability.

In both domains the adaptability is dependent on the past diagnostic result, monitoring goals and patient-specific features.

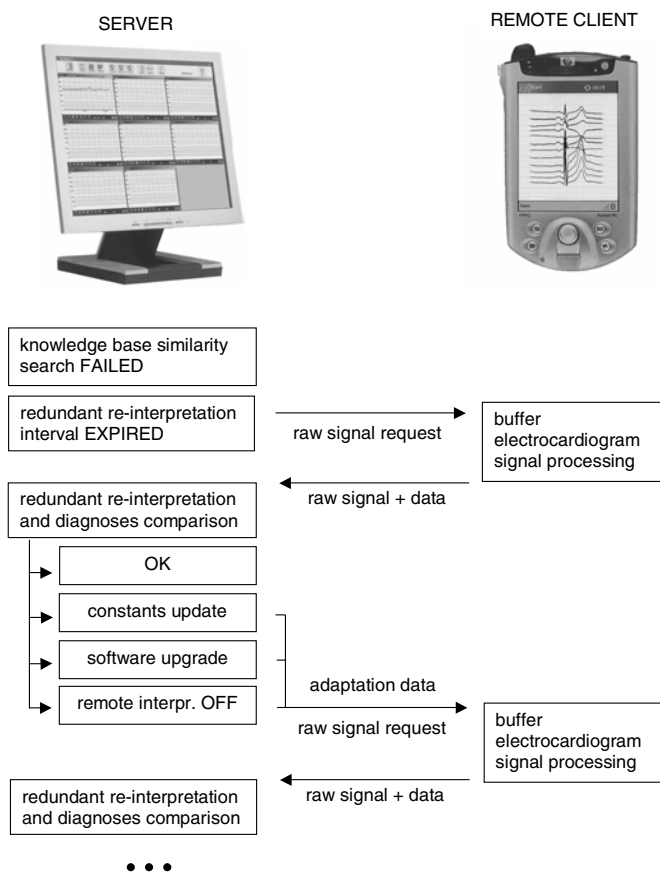


Fig. 1. The architecture of remote ECG recorder interpretation optimized for the data flow and processing error propagation

2.2 Concept of Multidimensional Quality Estimate

While the conventional rigid software has to be evaluated in the domain of result accuracy only, the adaptive software may be assessed in the multidimensional hyperspace including:

- asymptotic accuracy,
- adaptation delay,
- convergence (adaptation correctness).

Asymptotic accuracy is the absolute value of diagnostic error when the transient-evoked software adaptation is completed. Assuming no other transient is present in the subsequent signal it may be expressed as:

$$Q = \lim_{t \rightarrow \infty} |v(t) - v_0| \quad (1)$$

where $v(t)$ is subsequent diagnostic outcome and v_0 is the absolute correct value. Adaptation delay is defined as the time period from the transient occurrence t_0 to the moment t_D when the diagnostic outcome altered by the interpreting software modification starts falling into a given tolerance margin ε around its final value.

$$D = t_D - t_0 : \forall t > t_D \ v(t) \in (v(\infty) - \varepsilon, v(\infty) + \varepsilon) \quad (2)$$

The convergence represents the correctness of decisions made by the management procedure about the interpretation processing chain. Taking the analogy from the theory of control, the software adaptation plays the role of a feedback correcting the diagnoses made automatically. If the software modification decisions are correct, the outcome altered by the interpreting software modification approaches to the true value, the modification request signal is removed in consequence of decreasing error and the system is stable. Incorrect decisions lead to the growth of diagnostic outcome error and imply even stronger request for modification. The outcome value may stabilize on an incorrect value or swing the measurement range in response to subsequent modification attempts. In such case the system is unstable and the diagnostic outcome does not converge to the true value.

2.3 Concept of Weighted Accuracy Estimate

There is no general estimate of diagnostic quality and in the system composed of several procedures responsible for each parameter, the quality estimates need to correspond to the procedures selected for modification. Usually in the ECG interpretation chain there is a complex dependence of the diagnostic parameters and interpreting procedures (fig. 2). Each final outcome is influenced by several procedures and each procedure usually affects multiple parameters. The range of influence depends on the interpretation stage at which the procedure is applied. The quality of early processing stages affects all diagnostic parameters and the influence range gets narrower at subsequent stages. Each kind of diagnostic procedure is attributed by a static list of influenced diagnostic parameters.

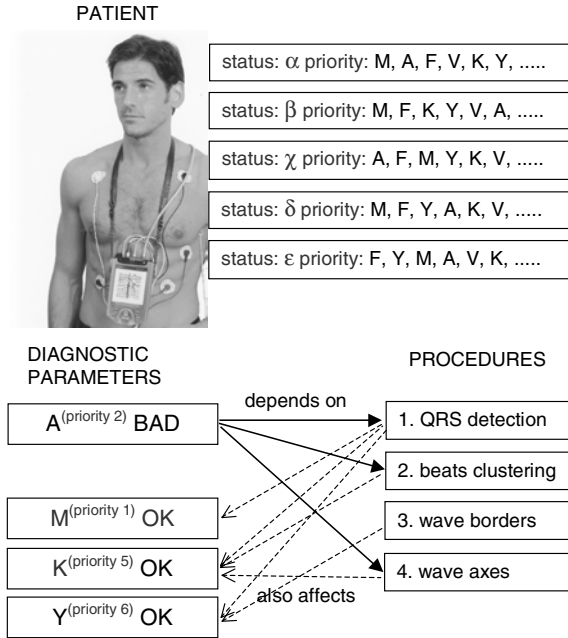


Fig. 2. The illustration of reciprocal dependencies of diagnostic parameters and interpretation procedures. Accordingly to the patient status, the parameters priority influences the final decision on remote software management. A bad parameter A triggers the replacement of only the procedures '2. beats clustering' and '4. wave axes' affecting parameters of lower priority K and Y).

The system makes its decision about the software modification with regard to all diagnostic parameters that may be concerned. The list of influenced diagnostic parameters is hierarchically scanned in order to detect any conflict of interest between simultaneously affected data. This hierarchy is, however, variable depending on patient status. Following the dependence of the diagnostic parameters medical relevance on the patient status, we propose the use of the same list of relevance factors to weight the contribution of particular parameters error to the general estimate of diagnostic quality.

2.4 Providing the Uniform Data

Non-uniform asynchronous updating of particular diagnostic parameters is an intrinsic advantage of adaptive interpretation systems, however direct comparison of their outcome to the reference values is not possible. The patient status has to be estimated from the irregular series of data issued by the adaptive system under test at each data point when the reference results are available. The diagnostic outcome of the adaptive interpretation being non-uniformly sampled time

series $N_j(n, v(n))$ was first uniformized with use of the cubic spline interpolation [1] given by a continuous function:

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \quad (3)$$

The interpolation yielded the uniform representation of each parameter by sampling the $S_i(x)$ at the time points m corresponding to the results of the fixed software:

$$N'_j(m) = \sum_m S_i(x) \cdot \delta(x - mT) \quad (4)$$

$x \in [x_i, x_{i+1}]$, $i \in \{0, 1, \dots, n-1\}$ best fitted to the series N_j . The values estimated at regularly distributed time points were finally compared to the reference. The assessment of data conformance has to consider three quality factors: tested data accuracy at their individual sampling points, interpolation error and reference data accuracy.

3 Results

The behavior of limited-scale network prototype of an ECG monitoring system with auto-adaptive software was investigated with use of the proposed tools. The remote recorder was based on a PDA-class handheld computer with ADC module (8 channels, 12 bits, 500 sps) and bi-directional GPRS connection. The stationary server was a PC-class desktop computer with a static IP address and 100Mb Internet access running Linux OS. The database contained 857 signals composed of artificially joined physiological ECG and a signal representing one of 14 pathologies under question. The main goal of the test was the assessment of the software adaptation correctness. The remote software update process is initiated if remote-issued diagnostic results differ from the server-calculated reference by more than a threshold defined accordingly to the diagnosis priority in four categories: 2% for QRS detection and heart rate, 5% for wave limits detection and ST-segment assessment for ischemia, 10% for morphology classification and 20% for remaining parameters. In case the result after a single software modification step is still out of the given tolerance margin, the decision about next update is made upon the new value is closer to the reference, and allow up to four consecutive update steps.

In technical aspect, the correctness of software upgrade and replacement is expressed by the percentage of incorrect adaptation attempts. As such were considered resources overestimation, leading to allocation violation, and underestimation, resulting in suspending of the software upgrade when the upgrade was feasible (tab. 2). In medical aspect, the correctness of interpretive software upgrade and replacement is expressed by the percentage of adaptation attempts leading to diagnostic parameters converging to the reference values (tab. 3). The overall distance in the diagnostic parameters hyperspace is expressed by the values of diagnostic parameters errors weighted by diagnosis priority.

The quality tests demonstrated for this particular system the technical correctness of 768 (89.6%) of adaptation attempts, among of them 643 (99.4%) software

Table 1. Results of remote diagnostic results convergence test after the consecutive steps of interpretation software modification

calculation constants	update steps	converging %	non-converging %
first		63.1	36.9
second		74.5	25.5
third		79.1	20.9
fourth		80.7	19.3

Table 2. Technical correctness of software upgrade and replacement

action	upgrade possible	upgrade impossible
software upgrade	647 (75.5%)	27 (3.1%)
library replacement	62 (7.3%)	121 (14.1%)

Table 3. Medical correctness of software adaptation

action	diagnosis improvement	diagnosis degradation
software upgrade	643 (99.4%)	4 (0.6%)
library replacement	97 (80.2%)	24 (19.8%)

upgrades and 97 (80.2%) software replacements yielded diagnostic results similar to the reference observed from the experts' survey (medical correctness). In 63.1% of cases the modification completed in a single iteration of 4.4 s average duration. The total of 89 (10.4%) adaptation attempts failed in result of incorrect estimation of resources availability. Resources overestimation resulting in the remote OS crash and thus monitoring discontinuity occurred in 27 (3.1%) cases.

4 Conclusions

Presented method offers estimations of various new features emerging due to the ECG processing adaptivity. Several concepts presented in the paper (multidimensional quality estimate, weighted accuracy estimate) reveal high complexity of the problem and some area not covered by the medical procedures and recommendations. These topics were presented for discussion on a cardiologist's forum. Principal elements of proposed quality estimation method was used for assessment of a prototype cardiac monitoring network. In this application our method contributed to final adjustment of the system properties in particular automatic decision making about further processing and reporting in a remote recorder.

Acknowledgement. Scientific work supported by the AGH-University of Science and Technology grant No 10.10.120.783.

References

1. Aldroubi, A., Feichtinger, H.: Exact iterative reconstruction algorithm for multivariate irregularly sampled functions in spline-like spaces: the L_p theory. Proc. Amer. Math. Soc. 126(9), 2677–2686 (1998)
2. Augustyniak, P.: The use of selected diagnostic parameters as a feedback modifying the ECG interpretation. Proc. Computers in Cardiology 33, 825–828 (2006)
3. IEC, 60601-2-47. Medical electrical equipment: Particular requirements for the safety, including essential performance, of ambulatory electrocardiographic systems (2001)
4. Willems, J.: Common Standards for Quantitative Electrocardiography 10-th CSE Progress Report. Leuven: ACCO publ. (1990)

Projective Versus Linear Filtering for Repolarization Duration Measurement

Marian Kotas

Institute of Electronics, Silesian University of Technology, 44-100 Gliwice,
ul. Akademicka 16
marian.kotas@polsl.pl

Summary. Automatic measurement of repolarization duration is highly prone to errors. The paper is focused on the problem of ECG noise suppression prior to the measurement. Two methods are compared: the traditional linear lowpass filtering and the projective filtering of the time aligned ECG beats. The performed experiments show that the projective filtering has much stronger impact on the precision of the measurements.

1 Introduction

The time between depolarization and repolarization of ventricles is covered by the QT interval. It is an important electrocardiographic parameter, often used to quantify the duration of ventricular repolarization. Since precise determination of a Q wave onset and a T wave end is relatively difficult, new variables were defined to quantify the repolarization duration (RD) in the research focused on the analysis of RD variability [1, 2]. It was shown [1] that the most precise measures can be obtained if the left limit of the repolarization interval is defined as the position of the R wave maximum and the right limit as the position of the T wave maximum. The RT_{max} interval duration is highly correlated with the duration of the actual QT interval [2]. However, it was reported [3] that the interval between the T wave peak and its end contains important information, which should not be discarded. Therefore this paper is focused on the measurements of RT_{end} interval. Since measurements of RT_{end} interval can be performed on the basis of high quality signals only, they should be preceded by an operation of ECG signal enhancement. In [4] it was shown that a significant reduction of the measurement errors can be achieved by application of projective filtering of time-aligned ECG beats (PFTAB). The goal of this paper is to investigate the PFTAB influence on the precision of the repolarization duration measurement and to compare the method to a classical linear filtering.

2 Methods

2.1 Projective Filtering of Time-Aligned ECG Beats

The method performs the following operations [5].

1. Reconstruction of the state-space representation of the observed noisy signal, by application of the Takens embedding operation, where a point in the constructed space is a vector:

$$\mathbf{x}^{(n)} = [x(n), x(n + \tau), \dots, x(n + (m - 1)\tau)]^T \quad (1)$$

where $x(n)$ is the processed signal, τ is the time lag ($\tau = 1$ is used in this application), m is the embedding dimension.

2. Correction of individual points $\mathbf{x}^{(n)}$ of the trajectory.
3. Conversion of the corrected trajectory points back into one-dimensional signal.

An outline of the algorithm of trajectory points correction (performed by locally linear projections, according to the rules of Principal Component Analysis – PCA [6]) is as follows.

After QRS complex detection, the points occupying the same position within the respective ECG beats are gathered in small neighborhoods. Within each neighborhood the local mean is computed and the covariance matrix of the points deviations from the mean. The matrix undergoes eigendecomposition. The eigenvectors associated with the assumed number q of the largest eigenvalues compose the q -dimensional principal subspace. The respective points contained in the analyzed neighborhoods are projected [6] into the constructed principal subspaces to suppress noise while preserving the desired component variability. Then one-dimensional representation of the processed signal is being reconstructed. More comprehensive description of the method can be found in [5].

2.2 Resampling the ECG Signal

Projective filtering can effectively be applied to processing the signals in low sampling rate [5], but determination of repolarization interval limits according to the algorithm which was applied in this study (developed by Laguna et. al. [7]) is of low accuracy in such conditions. In order to raise this accuracy the signals obtained after projective filtering can be resampled with higher frequency. In [1] it was shown (and our experiments confirmed it) that this operation raises the accuracy of repolarization duration measurement.

2.3 Low Pass Differentiator for Repolarization Duration Measurement

The algorithm described in [7] belongs to the classic approaches to QT interval measurement. First, the signals are prefiltered by a cascade of a digital differentiator and a moving average filter. The filters were developed for the sampling

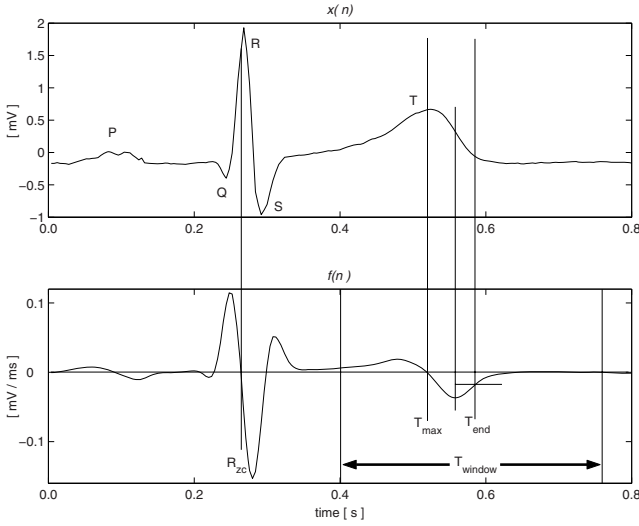


Fig. 1. The analyzed ECG signal ($x(n)$) and the output ($f(n)$) of the cascade (2). T_{window} is the region of the search for T_{max} and T_{end} .

frequency of 250 Hz. If the actual sampling frequency of the analyzed signal is L times higher the transfer function of the cascade can be adjusted to this frequency as follows

$$H(z) = (1 - z^{-6 \cdot L}) \left(\frac{1 - z^{-8 \cdot L}}{1 - z^{-1}} \right). \quad (2)$$

In the original algorithm the output of the differentiator is analyzed to determine the QRS onset. This operation is a source of relatively high errors. Fortunately, it is not crucial if RD variability is concerned [2]. Therefore in this study we employ only that part of the algorithm which performs the analysis of the cascade output (denoted as $f(n)$ in Fig. 1). The left limit of the repolarization interval is defined as a point where $f(n)$ crosses zero between the two extreme peaks which correspond to the highest ascending and descending slopes of the R wave. This point will be denoted as R_{zc} (zc stands for zero crossing). The right limit is either the T wave maximum (or minimum if the wave is inverted) or the T wave end. T wave maximum (T_{max}) is similarly a point where $f(n)$ crosses zero between the two extreme peaks in the automatically established [7] search window (T_{window} in Fig. 1). To establish the T wave end (T_{end}) we first search for the minimum of $f(n)$ after T_{max} (this minimum corresponds to the highest slope after T_{max}). Then we search for the point where $f(n)$ crosses a threshold which is equal to the half of this minimum. Details of the algorithm (and its action when T wave has different shape) can be found in [7].

3 Numerical Experiments and Discussion

Six signals of high quality from the QT database [8] (stored with the sampling frequency of 250 Hz) contaminated additively with either white Gaussian or real electromyographic (EMG) noise were used to test the method's performance. We performed R_{zc} , T_{max} and T_{end} determination in the "noise free" signals, in the contaminated ones, and in the contaminated signals enhanced either by projective or by linear filtering. On this basis we calculated three types of errors.

- e_1 - calculated as the difference between the position determined in the contaminated signal and the corresponding "true" position in the noise free signal. This type of error is related with the accuracy of the original procedure of R_{zc} , T_{max} and T_{end} determination.
- e_2 - calculated as the difference between the position determined in the signal obtained after projective filtering and the corresponding "true" position. This type of error shows the influence of projective filtering on the accuracy of R_{zc} , T_{max} and T_{end} determination.
- e_3 - calculated as the difference between the position determined in the signal obtained after linear low-pass filtering and the corresponding "true" position. This type of error shows the influence of classical filtering on the accuracy of R_{zc} , T_{max} and T_{end} determination.

During experiments the following parameters of the projective filter were applied: $m=50$, $q=3$. The classical low pass filtering was performed by application of Butterworth filters of various order and various cut-off frequency. To avoid phase distortion double filtering was performed: in forward and reverse direction. Fig. 2 illustrates the filter parameters influence on T_{end} determination errors in white noise environment. We can notice that for different filter orders similar

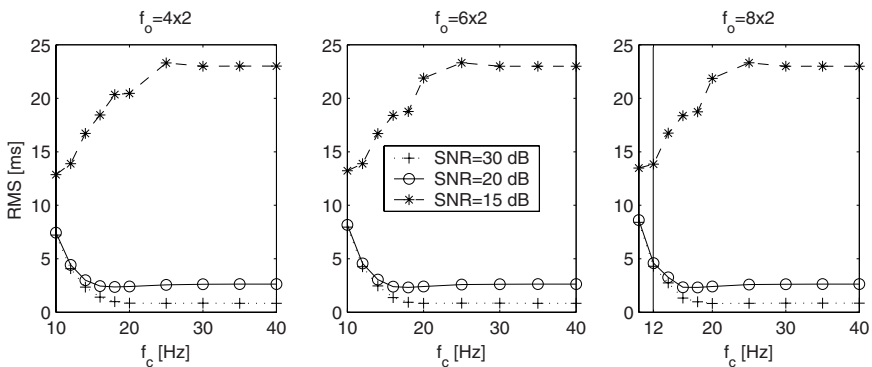


Fig. 2. Root mean square of the T_{end} determination error e_3 as a function of the cut-off frequency (f_c) of the applied low-pass Butterworth filter. The respective panels correspond to different filter orders (f_o); the multiplication by 2 resembles that double filtering in forward and reverse direction was performed.

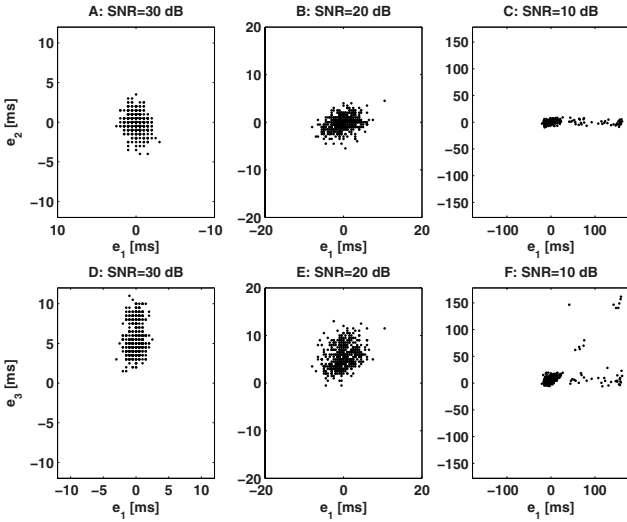


Fig. 3. Accuracy of T_{end} determination in white noise environment: distribution of points in the e_2 versus e_1 (the upper panels) or e_3 versus e_1 plane

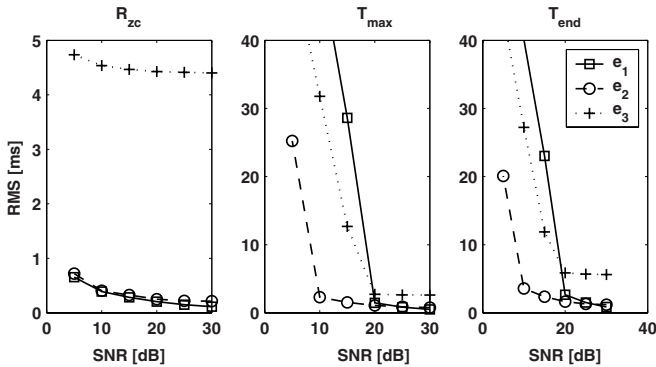


Fig. 4. The RMS value of the errors e_1 e_2 e_3 as the function of the SNR

results were obtained. The filter cut-off frequency has more evident influence on e_3 errors. For SNR=15 dB the errors fall with the decrease of f_c . However, a too low value of this parameter results in the desired ECG deformation which is signalled by the increase of the error with the decrease of f_c for lower level of noise (SNR=30, and SNR=20 dB). For further experiments the filter with $f_o=8x2$ and $f_c=12$ Hz was chosen.

The influence of either projective or linear filtering on T_{end} determination errors is illustrated in Fig. 3. We can notice that projective filtering significantly decreased the errors for SNR=20 and for SNR=10 dB (panels B and C).

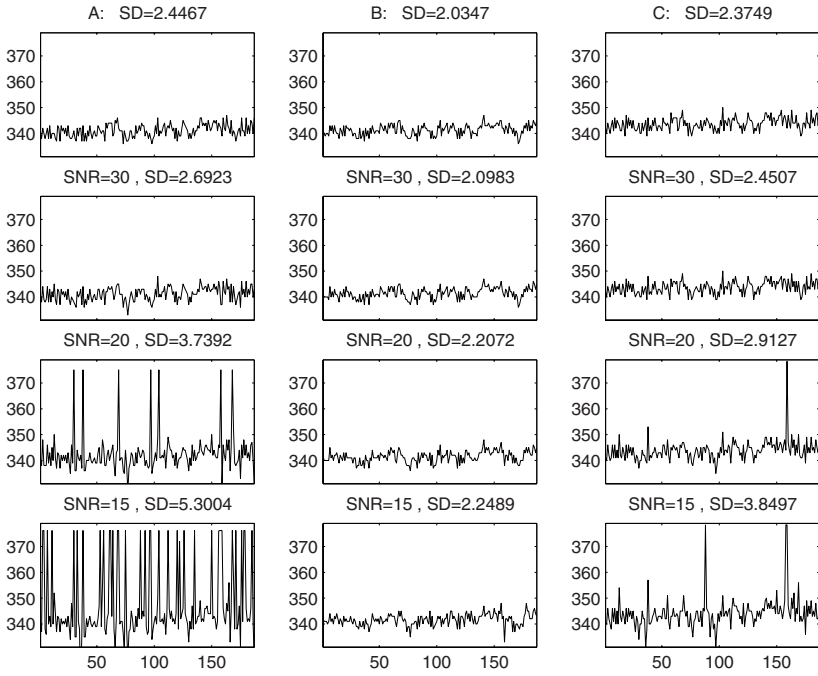


Fig. 5. The time series of RT_{end} intervals. Each series is characterized by its standard deviation (SD). The uppermost series (A,B,C) were established on the basis of a high quality ECG signal from the QT database. Below are the results obtained for the signal contaminated additively with EMG noise. The left column corresponds to the original measuring procedure, the middle one shows the results of projective filtering and the right one the results of linear filtering.

Moreover, projective filtering prevented the measuring procedure from a complete failure which is signalled by large errors. Such errors occurred for SNR=10 dB when the measurements were not preceded by projective filtering.

Linear low-pass filtering with the very low cut-off frequency $f_c=12$ Hz also decreased the T_{end} determination errors for SNR=10 and for SNR=20 dB. However, for SNR=30 dB (panel D) it caused a systematic error (we can notice only positive values of e_3). It explains the growth of the root mean square of this error with the decrease of f_c for low level of noise (in Fig. 2). Applying this type of filtering to achieve better repeatability of the measurements for higher level of noise, we have to accept this bias.

All the distributions obtained for the respective characteristic positions (R_{zc} , T_{max} and T_{end}) were evaluated by the RMS values of the three types of errors. The results are presented in Fig. 4. We can notice that projective filtering caused a significant reduction of errors for T_{max} and T_{end} determination while preserving the accuracy of R_{zc} determination. The classical filtering, on the other hand,

helped raise the repeatability of T_{max} and T_{end} determination but it introduced a high bias to these measurements (which is signalled by a higher level of the RMS error for high SNR) and a very high bias to R_{zc} determination.

The time series established on the basis of a high quality signal from the QT database are presented in Fig. 5. We can notice that the measurements are highly prone to errors. Even the noise of very low level (SNR=30 dB) results in measurement errors that raise the standard deviation of the time series calculated by the original procedure. For higher level of noise (SNR=20 dB) the series contains many spikes that result from a complete failure of the procedure. Linear pre-filtering prior to the measurements decreased the number of large measurement errors and made the procedure be slightly more immune to the noise contaminating the desired signal. However, for SNR=15 dB the measurement errors significantly raised the standard deviation of the RT_{end} series. When projective filtering was applied for ECG signal enhancement prior to the measurements, the procedure was prevented from large errors. Moreover, the series obtained for different levels of added noise had very similar statistical properties (as measured by standard deviation). We can conclude that for SNR higher than 15 dB the method allows to obtain very precise measurements which makes possible analysis of the repolarization duration statistical properties.

4 Conclusions

The modern approaches to patients risk stratification on the basis of the ventricular repolarization analysis meet different difficulties. Among the most important are the measurement errors. Particularly difficult is the beat-to-beat analysis of the ECG signal features. The necessity to preserve details of the dynamically changing morphology of the successive beats limits the possibility of applying the classical time averaging prior to the measurements. Nonlinear projective filtering, the method which allows ECG noise suppression with limited reduction of the desired signal morphological variability, appeared very useful in this application. Applied to ECG enhancement prior to the measurements of the QT interval, it significantly raised the measurements accuracy. Compared to the classical linear lowpass filtering, the method was much more effective. It allowed to obtain the repeatable measurements in a wide range of the signal-to-noise ratio. More precise measurements make possible more reliable analysis of the repolarization duration variability.

References

1. Tikkanen, P.E., Sellin, L.C., Kinnunen, H.O., Huikuri, H.V.: Using simulated noise to define optimal QT intervals for computer analysis of ambulatory ECG. *Med. Eng. Physics* 21, 15–25 (1999)
2. Merri, M., Alberti, M., Moss, A.J.: Dynamic Analysis of Ventricular Repolarization Duration from 24-hour Holter recordings. *IEEE Trans. Biomed. Eng.* 40, 1219–1225 (1993)

3. Davey, P.P.: QT interval measurements: Q to T_{Apex} or Q to T_{End}. J. Intern. Med. 246, 145–149 (1999)
4. Kotas, M.: Projective filtering of the time-aligned ECG beats for repolarization duration measurement. Computer Methods and Programs in Biomedicine 85, 115–123 (2007)
5. Kotas, M.: Projective Filtering of the Time-Aligned ECG Beats. IEEE Trans. Biomed. Eng. 51, 1129–1139 (2004)
6. Jolliffe, I.T.: Principal Component Analysis. Springer, New York (1986)
7. Laguna, P., Thakor, N.V., Caminal, P., Jane, R., Yoon, H.: New Algorithm for QT Interval Analysis in 24 Hour Holter ECG: Performance and Applications. Med. Biol. Eng. Comput. 28, 67–73 (1990)
8. Laguna, P., Mark, R.G., Goldberg, A., Moody, G.B.: A Database for Evaluation of Algorithms for Measurement of QT and Other Waveform Intervals in the ECG. Computers in Cardiology 24, 673–676 (1997)

An Application of Robust Kernel-Based Filtering of Biomedical Signals

Tomasz Pander

Silesian University of Technology,
Institute of Electronics, Division of Biomedical Electronics,
Akademicka 16, 44-100 Gliwice, Poland
tpander@polsl.pl

Summary. Noises which appear during recording of biomedical signals are seldom characterized by the ideal Gaussian model, because noises can have an impulsive nature. In this work the robust filter is introduced on the base of the Gaussian radial basis function. For that purpose the distance measure in the feature space is defined. Minimizing of this distance measure allows to estimate output value in a moving window filter. Presented filter was applied to suppress muscle noise in ECG signal. Some properties of presented filter are also shown. The experiments performed show that introduced filter is effective in suppression of noise of biomedical signals.

1 Introduction

The process of a noise suppression should be the first step of each biomedical signal processing system. The accuracy of further signal processing steps like detection, measurement or classification depend on the quality of reduction noise algorithms [5]. There exists many different biomedical signals, but for the purpose of this work the electrocardiogram (ECG signal) was chosen. Electrocardiographic (ECG) signals are composed of series of component waves (P, QRS and T waves) separated by iso-electric regions. There are the following types of noise that disturbed the ECG signal during a stress test: the baseline wander noise, electrode motion artifact and electromyogram-induced noise (EMG) [1]. The muscle noise contaminations in ECG signals distort low-amplitude ECG wave components and hence lower the accuracy of computer-aided measurements of various morphological characteristics [3]. The most popular model of EMG noise is a Gaussian model, but the muscle noise shows frequently an impulsive nature. Another model which describes very likely a muscle noise is model based on the symmetric α -stable distributions [7]. One of the most disadvantage of linear filters are their sensitivity to outliers samples, which can appear as an impulsive noise or sudden change of signal's morphology. At such condition frequently applied a group of linear filters may fail. More appropriate should be a group of non-linear filters which are robust to outliers in a filtered signal.

The main purpose of this paper is to present a non-linear robust filter which is based on the kernel function. The Gaussian radial basis function is applied as the kernel function. The usefulness of the presented filter is shown in obtained

results. The rest of this paper is organized in the following way. The next section presents the theory of kernel methods. In the section 3, the robust kernel-based filter is presented. The section 4 presents the method of the evaluation of the presented filter and some results. Final conclusions are presented in the last section.

2 Kernel Methods

Kernel-based algorithms have been recently developed in the machine learning community. The general idea of the kernel methods is the following [6, 8]. Given a linear algorithm, one first maps the data living in a space \mathcal{X} (the input space) to a vector space \mathcal{H} (the feature space) via a nonlinear transformation $\phi(\cdot) : \mathcal{X} \rightarrow \mathcal{H}$, and then runs the algorithm on the data representation $\phi(\mathbf{x})$ of the data. Above description can be written in the following way [9]:

$$\phi : \mathbf{x} \in X \subseteq R^n \rightarrow \phi(\mathbf{x}) \in H \subseteq R^N \quad (n \ll N). \quad (1)$$

The form of $\phi(\cdot)$ can be hard to guess. If only inner products between data vectors are considered, the data appears in expressions like

$$\langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle = \phi^T(\mathbf{x})\phi(\mathbf{y}). \quad (2)$$

Here can be used the fact that for a certain specific nonlinear transformation $\phi(\cdot)$ the inner product can be directly computed from the data \mathbf{x} and \mathbf{y} without computing $\phi(\mathbf{x})$ and $\phi(\mathbf{y})$ [8]. This is known as the kernel function. The kernel function k is defined as $k(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle$ [4, 8]. A kernel k is a symmetric function of two variables. The most popular is the family of radial basis kernel functions (RBF) which can be expressed as $k(\mathbf{x}, \mathbf{y}) = r(\|\mathbf{x} - \mathbf{y}\|)$, where $r(\cdot)$ is an arbitrary function, like the Gaussian kernel which has the following form [6, 10]:

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{-\|\mathbf{x} - \mathbf{y}\|^2}{\sigma^2}\right), \quad (3)$$

where σ^2 is the variance parameter (or the kernel's width). For all RBF kernels $\forall_x k(\mathbf{x}, \mathbf{x}) \equiv 1$ [9]. More information about kernel functions can be found in [2, 4, 8, 10, 11].

3 Kernel-Robust Filter

Let us consider the desired signal s_i which is disturbed with noise components v_i , where i is the discrete time index. Then the input (recorded) signal x_i can be written as:

$$x_i = s_i + v_i. \quad (4)$$

The main aim of filtering is to estimate the signal samples s_i by using the noisy samples x_i . The output of the filter is $y_i = \mathbf{F}(x_i)$, where $\mathbf{F}(\cdot)$ denotes the

filtering operation. Let $d(\mathbf{x}, \mathbf{y})$ denotes the distance measure in the feature space between signals x_i and y_i . Using the RBF kernel, the distance measure $d(\mathbf{x}, \mathbf{y})$ is defined as:

$$\begin{aligned} d(\mathbf{x}, \mathbf{y})^2 &= \|\phi(\mathbf{x}) - \phi(\mathbf{y})\|^2 = \phi(\mathbf{x})^T \phi(\mathbf{x}) - 2\phi(\mathbf{x})^T \phi(\mathbf{y}) + \phi(\mathbf{y})^T \phi(\mathbf{y}) \\ &= k(\mathbf{x}, \mathbf{x}) - 2k(\mathbf{x}, \mathbf{y}) + k(\mathbf{y}, \mathbf{y}) = 2 - 2k(\mathbf{x}, \mathbf{y}). \end{aligned} \tag{5}$$

This distance measure $d(\mathbf{x}, \mathbf{y})$ corresponds exactly to a class of new non-Euclidian distances in the original space with varying kernels and is robust to outliers [2, 9]. Considering the Gaussian RBF kernel's width σ^2 so large that $\sum_{i=1}^n (x_i - y_i)^2 \ll \sigma^2$ and from the fact that $1 - \exp(-x) \leq x$, then the Gaussian RBF kernel function can be written as [9]:

$$k(\mathbf{x}, \mathbf{y}) \cong 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sigma^2}, \tag{6}$$

and replacing $k(\mathbf{x}, \mathbf{y})$ in (5) then it yields:

$$d(\mathbf{x}, \mathbf{y})^2 \cong \frac{2 \sum_{i=1}^n (x_i - y_i)^2}{\sigma^2}. \tag{7}$$

Presented in eq. (7) distance measure is similar to L_p -norm measures. On the base of this measure the robust filter is build. This robust filter is a running window filter with the length N . The output of this new filter in a discrete time i is defined as:

$$\hat{y}_i = \arg \min_{\beta} G(\beta, \mathbf{x}), \tag{8}$$

where $G(\beta, \mathbf{x}) = \sum_{i=1}^N d(x_i, \beta)^2$.

The filter output is one of the roots of equation $\frac{\partial G(\beta, \mathbf{x})}{\partial \beta} = 0$. For the simplification the Gaussian RBF is considered. Using (5) the solution can be written as:

$$\frac{\partial d(x_i, \beta)^2}{\partial \beta} = \frac{\partial (2 - 2k(x_i, \beta))}{\partial \beta} = -2 \frac{\partial k(x_i, \beta)}{\partial \beta} = 0, \tag{9}$$

then

$$\frac{\partial k(x_i, \beta)}{\partial \beta} = \frac{\partial}{\partial \beta} \left(\exp \left(-\frac{(x_i - \beta)^2}{\sigma^2} \right) \right) = \frac{2}{\sigma^2} (x_i - \beta) \exp \left(-\frac{(x_i - \beta)^2}{\sigma^2} \right). \tag{10}$$

Now the first derivative of $G(\beta, \mathbf{x})$ has the following form:

$$\frac{\partial G(\beta, \mathbf{x})}{\partial \beta} \propto \sum_{i=1}^N (x_i - \beta) \exp \left(-\frac{(x_i - \beta)^2}{\sigma^2} \right). \tag{11}$$

The solution of the last equation is the following:

$$\beta = \frac{\sum_{i=1}^N x_i \exp \left(-\frac{(x_i - \beta)^2}{\sigma^2} \right)}{\exp \left(-\frac{(x_i - \beta)^2}{\sigma^2} \right)} = \frac{\sum_{i=1}^N k(x_i, \beta) x_i}{\sum_{i=1}^N k(x_i, \beta)} = f(\beta). \tag{12}$$

Because eq. (12) is the function of β the iterative method is applied to estimated value of β . The initial value of β is the median value from the set of $\{x_i\}_{i=1}^N$. The last equation allows to estimate value of the presented GK filter's output at the moment i . The obtained solution is more general and different types of kernel functions can be applied. In this work the Gaussian RBF is considered. This filter is denoted as the GK filter.

4 Numerical Experiments

Filtering a signal in the time-domain should result in decreasing of unwanted components of the input signal. The filtering process shouldn't deform the signal, but there exists a group of filters which may introduce inadmissible deformations of the signal. The nonlinear filters belong to this group. For that reasons, the presented filter is evaluated using the normalized mean square error (NMSE) defined as $NMSE = \frac{\sum_{i=1}^L (y_i - s_i)^2}{\sum_{i=1}^L (s_i)^2} \cdot 100\%$, where: L - is the length of signal, y_i is the output of the evaluated filter, s_i is the deterministic part of a signal without a noise. The NMSE factor allows to measure the relative power of the additive residual distortions introduced by the nonlinear filtering. Signals y_i and s_i are aligned and they have the same time index.

For the testing purposes, high resolution ECG signal cycle (with very high SNR and sampled at 2 kHz) is used as the deterministic component with added different realization of random noise. An example of such ECG signal is presented in the Fig. 1A. Then the noise samples are added to ECG cycles with the known value of standard SNR factor (-5, 0, 5, 10 and 20 dB) and then such signal is filtered. In this work a simulated noise (a white Gaussian noise) and a real EMG signal samples (sampled at 2 kHz) are used. For each value of SNR, the simulations are repeated 100 times. On the base of the output filter signal, value of the NMSE factor is estimated. After it the averaged $NMSE_{ave}$ value is calculated and presented in Table 1. The reference filters are the moving average (MA) filter and the median (med) filter. All experiments were run in the MATLAB environment.

4.1 A Width of the GK Filter and a Width σ^2 of RBF Function

One of the main aspects of filtering is proper choice of the window length N . The filter window length determines a speed of filtering. In the Fig. 2 the changes of the window length N is presented in the function of the signal-to-noise ratio. This parameter N is estimated in order to obtain the smallest value of NMSE factor when the kernel width is $\sigma^2 = 1$.

For low values of SNR, the filter length should be larger then for higher values of SNR. In the Fig. 2 the difference between a real muscle noise and a Gaussian model is shown, especially for low value of SNR. But when SNR is higher, the differences between noises are smaller. When signal is corrupted with a Gaussian noise the window length of the GK filter can be shorter than in the case of a

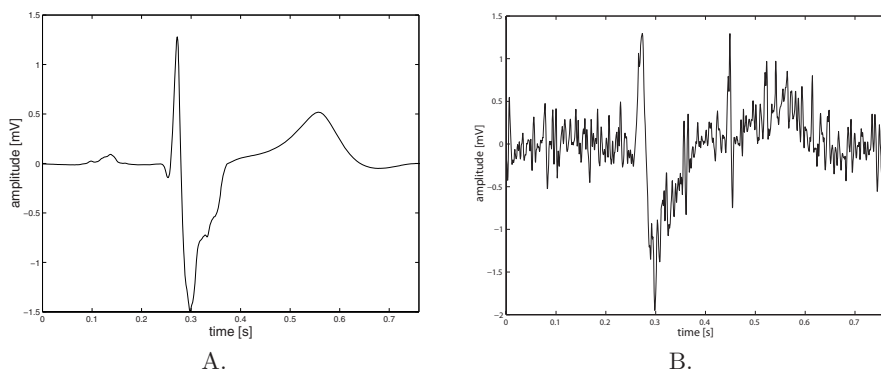


Fig. 1. The example of "clean" (A) and noisy (B) ECG signal (SNR=5 dB)

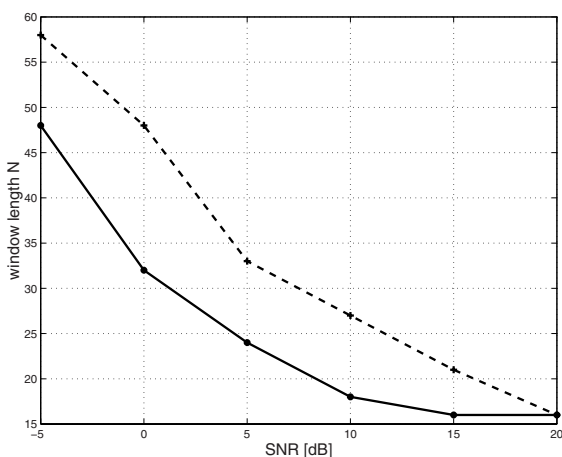


Fig. 2. The width of the GK filter N in the function of SNR level for ECG signal cycle corrupted with Gaussian noise (solid line) and EMG noise (dashed line)

real muscle noise. This parameter depends on a level of a noise. In this work for simplicity the length of tested filters is set to $N = 32$.

The second important parameter of presented filters is the kernel width σ^2 . In this paper two cases are considered. In the first case σ^2 is set to be 1. In the second case σ^2 value is calculated to reach the minimum value of the NMSE factor, viz. $\sigma_{opt}^2 \equiv \min(\text{NMSE}) |_{\sigma^2 = \sigma_{opt}^2}$. For that purpose the testing set of noises were created in similar way to the set of tested signals with noise. Obtained results are presented in the Fig.3. When SNR is lower ($\text{SNR} \leq 5$ dB), then σ^2 should be greater than 1. And when $\text{SNR} \geq 10$ dB the best results are obtained for $\sigma^2 < 1$. The differences between real muscle noise and Gaussian noise are small in the whole range of SNR.

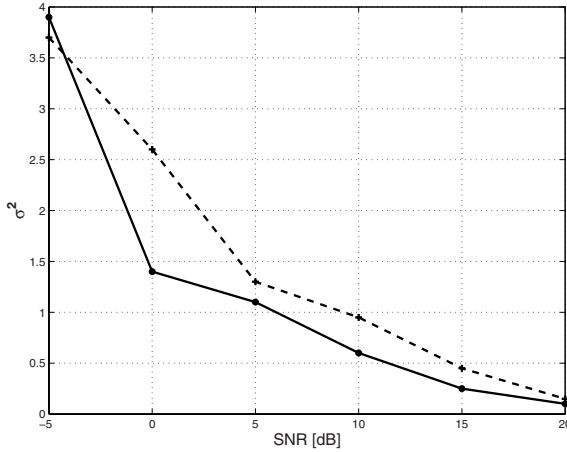


Fig. 3. The kernel width σ^2 as the function of SNR for the proposed GK filter for ECG corrupted with Gaussian noise (solid line) and real muscle noise (dashed line)

4.2 Signals with Gaussian and Real Muscle Noise

The aim of this experiment is to investigate the behavior of the GK filter in a presence of a simulated Gaussian noise and a real-life muscle noise. The first case is the ideal, because in practice a distribution of noise is more complicated. The second case is practical. Results are presented in the Table 1.

Table 1. Average Value of NMSE For ECG Signal Cycles Corrupted With Gaussian and Real Muscle Noise

SNR [dB]	White Gaussian Noise				Real Muscle Noise			
	MA	med	GK $\sigma^2 = 1$	GK σ_{opt}^2	MA	med	GK $\sigma^2 = 1$	GK σ_{opt}^2
-5	11.2245	16.2452	15.8651	11.4067	43.1992	54.6194	63.7873	44.6974
0	4.738	5.9036	4.6514	4.4256	14.9034	17.8359	16.727	15.1462
5	2.7358	2.4156	2.1703	2.1151	5.543	6.2683	5.6469	5.1807
10	2.1128	1.3371	1.4828	1.4577	2.9183	2.4046	2.4408	2.3079
20	1.8383	0.76593	1.167	1.0458	1.9336	0.93234	1.28	1.1958

Results presented in the Table 1 are obtained for filter window length $N = 32$. The left part of the Table 1 contains results obtained for a white Gaussian noise. The minimum value of NMSE factor was considered. When SNR= -5 dB, the best results of filtering are obtained for the MA filter. When SNR is in the range from 0 dB to 10 dB, presented GK filter leads to best results of filtering. For SNR=20 dB, the MA filter again leads to obtain the poorest results of filtering. Better results are obtained with proposed filter, especially when the parameter σ^2 is chosen optimally (minimized value of NMSE). But the smallest distortions

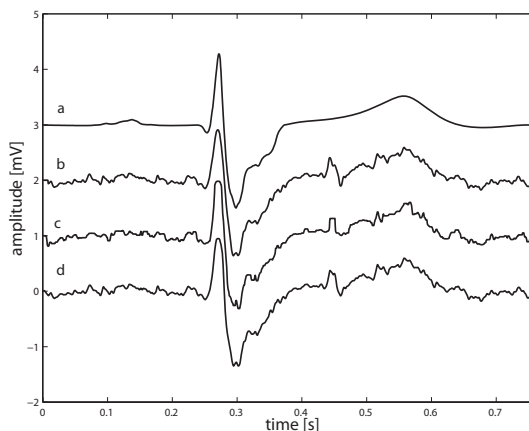


Fig. 4. The example of filtering of noisy ECG signal cycle (Fig. 1) and results of filtering (a - clean ECG cycle; b - MA (NMSE=4.37), c - med (NMSE=4.57), d - GK (NMSE=3.86) filters)

are introduced with the median filter. An estimation of the optimal σ^2 for GK filter improves the process of filtering for all cases of SNR. Results obtained for a muscle noise are presented in the right part of the Table 1. This type of noise has very often an impulsive nature. Results obtained for SNR=-5 dB are very poor for all tested filters. For SNR=0 dB and 5 dB the best results are obtained for the MA filter. When SNR varies from 5 to 10 dB, proposed GK filter leads to obtain the smallest value of NMSE factor for all cases ($\sigma^2 = 1$ and σ_{opt}^2). When SNR=20 dB the smallest NSME value reaches the median filter.

The example of filtering of the noisy ECG signal cycle (Fig. 1B) with using the proposed GK filter is shown in the Fig. 4.

5 Conclusions

The paper deals with the problem of a robust ECG signal processing and application of kernel functions for the purpose of filtering. Biomedical signals are usually contaminated by disturbances, which can have an impulsive nature. In this work the robust kernel-based filter is presented. The Gaussian RBF was chosen from many kernel functions. The principle of the presented filter is a process of minimizing the distance measure which is defined in the higher feature space. The necessary conditions for obtaining the minimum value of residual distortions are shown. It is the proper choice of the filter length and the RBF kernel function width σ^2 . A comparative study of the presented filter is included. The numerical examples show that filter based on the Gaussian RBF function leads to obtain comparative results of filtering to reference filters. In a few cases the GK filter leads to obtain the best results. Finally presented filter can be useful in ECG and other biomedical signals enhancement.

References

1. Alfonso, V., Tompkins, W., Nguyen, T., Michler, K., Luo, S.: Comparing Stress ECG Enhancement Algorithms. *IEEE Eng. in Medicine and Biology* 5, 37–44 (1996)
2. Zhang, D.-Q., Chen, S.-C.: Clustering Incomplete Data Using Kernel-Based Fuzzy C-means Algorithm. *Neural Processing Letters* 18, 155–162 (2003)
3. Hu, X., Nenov, V.: A single-lead ECG enhancement algorithm using a regularized data-driven filter. *IEEE Trans. on Biomedical Eng.* 2, 347–351 (2006)
4. Ch, J.-H., Pei-Yi, H.: A New Kernel-Based Fuzzy Clustering Approach: Support Vector Clustering With Cell Growing. *IEEE Trans. on Fuzzy Systems* 4, 518–527 (2003)
5. Leški, J.: Robust Weighted Averaging. *IEEE Trans. on Biomedical Engineering* 8, 796–804 (2002)
6. Müller, K.-R., Mika, S., Rätsch, G., Tsuda, K., Schölkopf, B.: An Introduction to Kernel-Based Learning Algorithms. *IEEE Trans. on Neural Networks* 2, 181–201 (2001)
7. Pander, T.: An application of a weighted myriad filter to suppression an impulsive type of noise in biomedical signals. *TASK Quarterly* 2, 199–216 (2004)
8. Pérez-Cruz, F., Bousquet, O.: Kernel Methods and Their Potential Use in Signal Processing. *IEEE Signal Processing Magazine*, 57–65 (May 2004)
9. Tan, K., Chen, S., Zhang, D.: Robust Image Denoising Using Kernel-Induced Measures, *Pattern Recognition*. In: *ICPR 2004, Proceedings of the 17th International Conference on Pattern Recognition* (2004)
10. Tan, Y., Wang, J.: A Support Vector Maching with a Hybrid Kernel and Minimal Vapnik-Chervonenkis Dimension. *IEEE Trans. on Knowledge and Data Engineering*, 385–395 (April 2004)
11. Vanschoenwinkel, B., Manderick, B.: Appropriate Kernel Functions for Support Vector Machine Learning with Sequences of Symbolic Data. In: Winkler, J.R., Niranjan, M., Lawrence, N.D. (eds.) *Deterministic and Statistical Methods in Machine Learning*. LNCS (LNAI), vol. 3635, pp. 255–279. Springer, Heidelberg (2005)

Weighted Averaging of ECG Signal Using Criterion Function Minimization

Alina Momot

Silesian University of Technology, Institute of Computer Science,
16 Akademicka St., 44-100 Gliwice, Poland
alina.momot@polsl.pl

Summary. Averaging signals in time domain is one of the main methods of noise attenuation in biomedical signal processing. This paper proposes a new weighted averaging method using criterion function minimization and based on partitioning of input set in time domain, which is a generalization of an existing method. Performance of the new method is experimentally compared with the traditional averaging by using arithmetic mean and another weighted averaging method based on criterion function minimization.

1 Introduction

Noise suppression plays very important role in the most of biomedical signal processing systems. Accuracy of all later operations performed on the signal, such as detections or classifications, depends on the quality of noise-reduction algorithms. Using the fact that certain biological systems produce repetitive patterns, an averaging in the time domain may be used for noise attenuation. Traditional averaging technique assumes the constancy of the noise power cycle-wise, however the most types of noise are not stationary. In reality it can be observed variability of noise power from cycle to cycle which is motivation for using methods of weighted averaging, which reduces influence of hardly distorted cycles on resulting averaged signal (or even eliminates them).

The electrocardiogram (ECG) is the recording of the heart's electrical potential versus time. The underlying physiological process, the electrochemical excitation of cardiac tissue, is non-linear and the signals show both fluctuations and remarkable structures which are not explained by linear correlations. These structures make the ECG useful to cardiologists as a diagnostic tool [6]. However, noise reduction is one of the main problems to overcome in order to obtain a correct interpretation of ECG signal and improvement the signal-to-noise ratio is usually very intractable because noise overlapping the signal in both time and frequency domains.

While the predominant QRS complexes which reflect the electrical depolarization of the ventricle are usually visible even in the presence of rather strong noise, more subtle features like the atrial P-wave (which normally occurs 120-200 ms

before the peak of the QRS complex) may be concealed by errors which are due to the imperfect transmission of the signal from the heart through different kinds of tissue, the electrode and the electronic equipment. These errors are particularly serious for ECG signal taken during exercise (sweaty skin, muscle activity) and on long-term ambulatory (Holter) recordings where the experimental conditions can not be controlled that well.

The peculiar twofold nature of ECG signal (a pronounced pattern repeated with irregular intervals) makes it difficult to filter these signals with Fourier methods. The continuous part of the spectrum is due to both the measurement noise (which should be removed) and to the irregular interbeat intervals (which should be preserved). However, using the fact of quasi-periodic nature of electrocardiographic signal, an averaging in the time domain may be used for noise attenuation and many recently developed noise removal techniques involve weighted signal averaging [1, 3, 4, 5].

The paper presents new method for resolving of signal averaging problem which incorporates partitioning of input set in time domain. By exploiting criterion function minimization it can be derived an algorithm of weighted averaging which application to electrocardiographic (ECG) signal averaging is competitive with alternative methods as will be shown in the later part of the paper. The new method is a generalization of the method proposed in [5], where the input set was divided into two separate subsets while the new method allows to divide the input set into any arbitrarily chosen number of disjoint subsets.

The paper is divided into four sections. Section 2 describes various methods of signal averaging, from the traditional arithmetic averaging, through the weighted averaging method WACFM [3], which will be treated as the reference method in numerical experiments, to proposed new weighted averaging method. Section 3 presents results of the numerical experiments. Conclusions are given in section 4.

2 Signal Averaging Methods

Let us assume that in each signal cycle $y_i(j)$ is the sum of a deterministic (useful) signal $x(j)$, which is the same in all cycles, and a random noise $n_i(j)$ with zero mean and variance for the i th cycle equal to σ_i^2 . Thus,

$$y_i(j) = x(j) + n_i(j), \quad (1)$$

where i is the cycle index $i \in \{1, 2, \dots, M\}$, and the j is the sample index in the single cycle $j \in \{1, 2, \dots, N\}$ (all cycles have the same length N). The weighted average is given by

$$v(j) = \sum_{i=1}^M w_i y_i(j), \quad (2)$$

where w_i is a weight for i th signal cycle ($i \in \{1, 2, \dots, M\}$) and $\mathbf{v} = [v(1), v(2), \dots, v(N)]$ is the averaged signal.

2.1 Traditional Arithmetic Averaging

The traditional ensemble averaging with arithmetic mean as the aggregation operation gives all the weights w_i equal to M^{-1} . If the noise variance is constant for all cycles, then these weights are optimal in the sense of minimizing the mean square error between \mathbf{v} and \mathbf{x} , assuming Gaussian distribution of noise. When the noise has a non-Gaussian distribution, the estimate (2) is not optimal, but it is still the best of all linear estimators of \mathbf{x} [2].

2.2 Weighted Averaging Method WACFM

In [3] it is presented algorithm WACFM (Weighted Averaging method based on Criterion Function Minimization). The idea of the algorithm is based on the fact that for $\mathbf{y}_i = [y_i(1), y_i(2), \dots, y_i(N)]^T$, $\mathbf{w} = [w_1, w_2, \dots, w_M]^T$ and $\mathbf{v} = [v(1), v(2), \dots, v(N)]$ minimization of the following scalar criterion function

$$I_m(\mathbf{w}, \mathbf{v}) = \sum_{i=1}^M (w_i)^m \rho(\mathbf{y}_i - \mathbf{v}), \tag{3}$$

where $\rho(\cdot)$ is a measure of dissimilarity for vector argument and $m \in (1, \infty)$ is a weighting exponent parameter, with respect to the weights vector yields

$$w_i = \frac{\rho(\mathbf{y}_i - \mathbf{v})^{\frac{1}{1-m}}}{\sum_{k=1}^M \rho(\mathbf{y}_k - \mathbf{v})^{\frac{1}{1-m}}}, \tag{4}$$

for $i \in \{1, 2, \dots, M\}$. When the quadratic function $\rho(\cdot) = \|\cdot\|_2^2$ is used, the averaged signal can be obtained as

$$\mathbf{v} = \frac{\sum_{i=1}^M (w_i)^m \mathbf{y}_i}{\sum_{i=1}^M (w_i)^m}, \tag{5}$$

for the weights vector given by (4) with the quadratic function. The optimal solution for minimization (3) with respect to \mathbf{w} and \mathbf{v} is a fixed point of (4) and (5) and it is obtained from the Picard iteration.

If parameter m tends to one, then the trivial solution is obtained where only one weight, corresponding to the signal cycle with the smallest dissimilarity to averaged signal, is equal to one. If m tends to infinity, then weights tend to M^{-1} for all i . Generally, a larger m results in a smaller influence of dissimilarity measures. The most common value of m is 2 which results in greater decrease of medium weights [3].

2.3 Proposed Weighted Averaging Method

In this section it is presented a new weighted averaging method using criterion function minimization and based on partitioning of input set in time domain. It is a generalization of method proposed in [5], where the input set was divided into two separate subsets while the new method allows to divide the input set into any arbitrarily chosen number of disjoint subsets.

As it is shown in [5], for input set $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M]$ (where $\mathbf{y}_i = [y_i(1), y_i(2), \dots, y_i(N)]^T$) divided into two disjoint subsets \mathbf{y}^1 and \mathbf{y}^2 , the weights vectors \mathbf{w}_1 and \mathbf{w}_2 are calculated as minimum of the following scalar criterion function

$$I(\mathbf{w}_1, \mathbf{w}_2) = (\mathbf{y}^1 \mathbf{w}_1 - \mathbf{y}^2 \mathbf{w}_2)^T (\mathbf{y}^1 \mathbf{w}_1 - \mathbf{y}^2 \mathbf{w}_2), \quad (6)$$

with constraints $\mathbf{w}_1 \mathbf{1} = 1$ and $\mathbf{w}_2 \mathbf{1} = 1$, which means that sum of weights for each vector is equal to one. Thus,

$$\mathbf{w}_1 = \frac{((\mathbf{y}^1)^T \mathbf{y}^1)^{-1} (\mathbf{y}^1)^T \mathbf{y}^2 \mathbf{w}_2 + \mathbf{1} - \mathbf{1}^T ((\mathbf{y}^1)^T \mathbf{y}^1)^{-1} (\mathbf{y}^1)^T \mathbf{y}^2 \mathbf{w}_2}{\mathbf{1}^T ((\mathbf{y}^1)^T \mathbf{y}^1)^{-1} \mathbf{1}} ((\mathbf{y}^1)^T \mathbf{y}^1)^{-1} \mathbf{1} \quad (7)$$

and

$$\mathbf{w}_2 = \frac{((\mathbf{y}^2)^T \mathbf{y}^2)^{-1} (\mathbf{y}^2)^T \mathbf{y}^1 \mathbf{w}_1 + \mathbf{1} - \mathbf{1}^T ((\mathbf{y}^2)^T \mathbf{y}^2)^{-1} (\mathbf{y}^2)^T \mathbf{y}^1 \mathbf{w}_1}{\mathbf{1}^T ((\mathbf{y}^2)^T \mathbf{y}^2)^{-1} \mathbf{1}} ((\mathbf{y}^2)^T \mathbf{y}^2)^{-1} \mathbf{1}. \quad (8)$$

In the method proposed in [5], there was made an assumption that input set is divided into two separate subsets. In the new method described below it can be done a division of the input set into any arbitrarily chosen number of disjoint subsets. The proposed new weighted averaging algorithm can be described as follows, where ε is a preset parameter:

1. Determine partitioning of input set into disjoint subsets \mathbf{y}^k of size M_k , where $k = 1, 2, \dots, K$ (K is at least equal 2) and $M_1 + M_2 + \dots + M_K = M$. Initialize weights $\mathbf{w}_K^{(0)}$. Set the iteration index $i = 1$.
2. Calculate $\mathbf{w}_1^{(i)}$ using (7), assuming $\mathbf{w}_2 = \mathbf{w}_K^{(i-1)}$ and $\mathbf{y}^2 = \mathbf{y}^K$. Calculate $\mathbf{w}_k^{(i)}$ as \mathbf{w}_2 using (8) for $k = 2, \dots, K$, assuming $\mathbf{y}^2 = \mathbf{y}^k$, $\mathbf{w}_1 = \mathbf{w}_{k-1}^{(i)}$ and $\mathbf{y}^1 = \mathbf{y}^{k-1}$.
3. If $\sum_{k=1}^K \|\mathbf{w}_k^{(i-1)} - \mathbf{w}_k^{(i)}\| > \varepsilon$ then $i \leftarrow i + 1$ and go to 2.
4. Calculate averaged signal $\mathbf{v} = \sum_{k=1}^K \alpha_k \mathbf{y}^k \mathbf{w}_k$, where $\alpha_k = M_k/M$.

3 Numerical Experiments

Performance of the new method was experimentally compared with the traditional averaging by using arithmetic mean and weighted averaging method based on criterion function minimization [3]. In all experiments, using weighted averaging, calculations were initialized as the means of disturbed signal cycles and

the parameter ε was equal to 10^{-6} . For a computed averaged signal the performance of tested methods was evaluated by the maximal absolute difference between the deterministic component and the averaged signal (MAX). The root mean-square error (RMSE) between the deterministic component and the averaged signal was also computed. All experiments were run in the R environment for R version 2.4.0 (www.r-project.org).

The simulated ECG signal cycles were obtained as the same deterministic component with added independent realizations of random noise. The deterministic component was ANE20000, taken from database CTS [7]. A series of 120 ECG cycles was generated with the same deterministic component and zero-mean white Gaussian noise with different standard deviations with constant amplitude of noise during each cycle. Figure 1 presents the ANE 20000 ECG signal (bold line) and this signal with Gaussian noise with standard deviation equal sample standard deviation of the deterministic component.

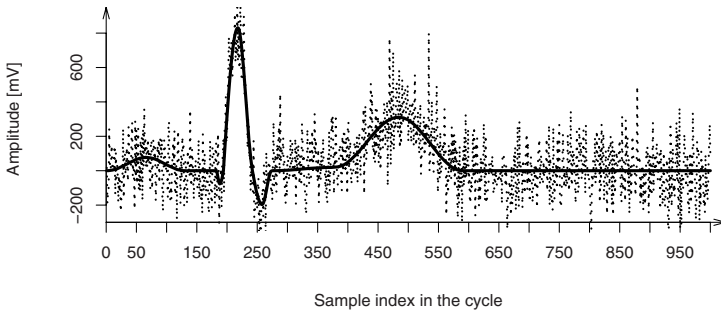


Fig. 1. The example of ECG signal and this signal with Gaussian noise

In consecutive experiments there were taken different noise amplitude characteristics varying from $0.05s$ to $2s$, where s is sample standard deviation of the deterministic component. In the first case the amplitude characteristic was step function of cycle index as depicted in Figure 2.

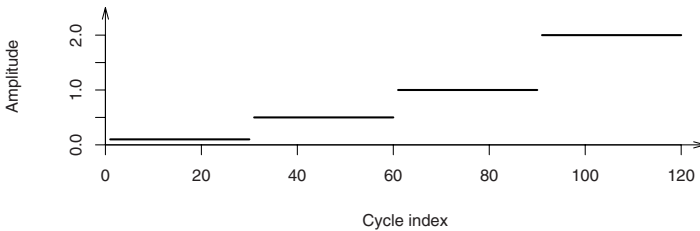


Fig. 2. Noise amplitude characteristic for the first case

Table 1. Results for the first case

Method	RMSE	MAX	NI
AA	15.10669	52.95812	–
WACFM	2.762457	9.085588	6
WAPM.2	2.825439	8.259491	15
WAPM.3	2.719333	8.151688	5
WAPM.4	2.704390	8.381335	4

Table 1 presents the RMSE and the absolute maximal value (MAX) of residual noise for all tested methods (the best results are bolded) such as traditional Arithmetic Averaging (AA), Weighted Averaging method based on Criterion Function Minimization (WACFM) with parameter m equal 2 and proposed new method of Weighted Averaging based on Partitioning of input set in time domain and criterion function Minimization (WAPM) with parameter $K = 2$ (WAPM.2), $K = 3$ (WAPM.3) and $K = 4$ (WAPM.4). For all methods, except of traditional Arithmetic Averaging, the number of iterations (NI) is also presented. Using presented algorithm the input set was divided into equal in number of elements subsets (i.e. $120/K$), where each of the subset indexes was equal to one plus remainder in division cycle index by K (interlaced partitioning, i.e. $\mathbf{y}^k = \{\mathbf{y}_k, \mathbf{y}_{k+K}, \mathbf{y}_{k+2K}, \dots, \mathbf{y}_{k+M-K}\}$ for $k = 1, 2, \dots, K$).

As can be seen the WAPM method with partitioning of input set into more than 2 subsets results in better performance. In this case the best results are given for WAPM.4 (division into 4 disjoint subsets) with respect to the root mean-square error (RMSE) and for WAPM.3 (division into 3 disjoint subsets) with respect to the maximal absolute difference between the deterministic component and the averaged signal (MAX). It is also worth noting that the number of iterations was significantly lower for the WAPM method with partitioning of input set into more than 2 subsets.

In the second case the amplitude characteristic was piecewise linear function of cycle index as depicted in Figure 3 and Table 2 presents results for this case. In this case also the WAPM method with partitioning of input set into more than 2 subsets results in better performance with respect to the root mean-square error (RMSE), but WAPM.2 (division into 2 disjoint subsets) gives the best result with respect to the maximal absolute difference between the deterministic component and the averaged signal (MAX). The number of iterations was also significantly lower for the WAPM method with partitioning of input set into more than 2 subsets.

In the third case the amplitude characteristic was another piecewise linear function of cycle index as depicted in Figure 4 and Table 3 presents results for this case. In this case, similarly as in the first case, the WAPM method with partitioning of input set into more than 2 subsets results in better performance both with respect to the root mean-square error (WAPM.3) and with respect to

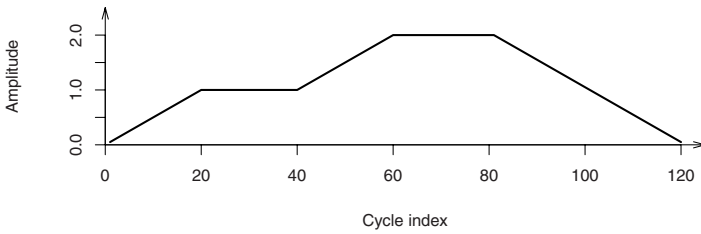


Fig. 3. Noise amplitude characteristic for the second case

Table 2. Results for the second case

Method	RMSE	MAX	NI
AA	17.49267	56.77732	–
WACFM	14.45609	51.60415	9
WAPM.2	6.671185	19.882892	14
WAPM.3	6.464871	22.062749	4
WAPM.4	6.535047	20.709526	4

the maximal absolute difference between the deterministic component and the averaged signal (WAPM.4) and the number of iterations was also significantly lower for the WAPM method with partitioning of input set into more than 2 subsets.

It is noticeable the relatively weak performance of WACFM algorithm in second and third case. In these cases often it was observed that the algorithm terminated unexpectedly without returning results. It may be the result of numerical unstability of its implementation.

There were also performed other experiments with different partitioning methods such as unequal in number of elements subsets and not interlaced partitioning, however the method with equal in number of elements subsets and interlaced partitioning resulted in best performance.

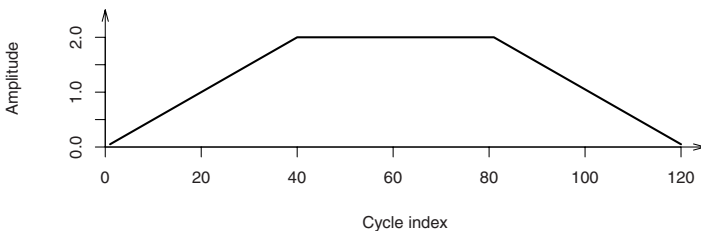


Fig. 4. Noise amplitude characteristic for the third case

Table 3. Results for the third case

Method	RMSE	MAX	NI
AA	20.31205	70.14208	–
WACFM	14.08565	40.40659	8
WAPM.2	6.61459	20.99080	14
WAPM.3	6.580835	23.013894	5
WAPM.4	6.596532	20.402297	4

4 Conclusion

In this work the new method of weighted averaging of biomedical signal was presented along with the application to averaging ECG signals. Presented new method, which incorporates partitioning of input set in time domain, resolves the weighted signal averaging problem and it is a generalization of the method proposed in [5]. By exploiting criterion function minimization it can be derived an algorithm of weighted averaging which application to electrocardiographic (ECG) signal averaging leads to results competitive with alternative methods such as weighted averaging method based on criterion function minimization proposed in [3].

References

1. Bataillou, E., Thierry, E., Rix, H., Meste, O.: Weighted averaging using adaptive estimation of the weights. *Signal Processing* 44, 51–66 (1995)
2. Łęski, J.: Application of time domain signal averaging and Kalman filtering for ECG noise reduction. Ph.D. Thesis, Silesian University of Technology, Gliwice (1989)
3. Łęski, J.: Robust Weighted Averaging. *IEEE Transactions on Biomedical Engineering* 49(8), 796–804 (2002)
4. Momot, A., Momot, M., Łęski, J.: Empirical Bayesian Averaging Method and its Application to Noise Reduction in ECG Signal. *Journal of Medical Informatics and Technologies* 10, 93–101 (2006)
5. Momot, A., Momot, M., Łęski, J.: Weighted Averaging of ECG Signals Based on Partition of Input Set in Time Domain. *Journal of Medical Informatics and Technologies* 11, 165–170 (2007)
6. Schamroth, L.: An introduction to electrocardiography. Oxford Press, Oxford (1986)
7. International Electrotechnical Commission Standard 60601-3-2 (1999)

Empirical Bayesian Approach to Weighted Averaging of ECG Signal Using Cauchy Distribution

Alina Momot¹ and Michał Momot²

¹ Silesian University of Technology, Institute of Computer Science,
16 Akademicka St., 44-101 Gliwice, Poland
alina.momot@polsl.pl

² Institute of Medical Technology and Equipment,
118 Roosevelt St., 41-800 Zabrze, Poland
michal.momot@itam.zabrze.pl

Summary. The analysis of the electrocardiographic signal recordings is greatly useful in the screening and diagnosing of cardiovascular diseases. However usually recording of the electrical activity of the heart is performed in the presence of noise. One of the commonly used techniques to extract a useful signal distorted by a noise is weighted averaging, since the nature of ECG signal is quasi-cyclic with level of noise power varying from cycle to cycle. This paper proposes a new weighted averaging method, which incorporates empirical Bayesian inference and the expectation-maximization technique. It is an extension of an existing method by introducing Cauchy distribution and the unknown parameter is estimated using interquartile range. Performance of the new method is experimentally compared with the traditional averaging by using arithmetic mean and other empirical Bayesian weighted averaging methods.

1 Introduction

An electrocardiogram (ECG) is the prime tool in non-invasive cardiac electrophysiology and has a prime function in the screening and diagnosing of cardiovascular diseases. In typical ECG signal, the dominant morphologies are the P, QRS and T waves. Occasionally a U-wave will be present immediately after the T-wave, the genesis of which is uncertain. The P-wave represents atria activation; the QRS complex represents ventricular activation or depolarization. An initial downward deflection after the P-wave is termed as 'Q', the dominant upward deflection is 'R' and the terminal part is denoted as 'S'. The T-wave represents ventricular recovery or depolarization. The ST segment, the T-wave and the U-wave together represent the total duration of ventricular recovery. The ST segment represents the greater part of ventricular depolarization. The ST segment usually merges smoothly and imperceptibly with the T-wave [4].

The electrocardiogram shows the results of nerve impulse stimuli by the heart, as the current is diffused around the surface of the body. The current at the body surface will be built on the voltage drop, which is a couple of V to mV with impulse variations. This is very small amplitude of impulse that needs to

be amplified to an amount, which is large enough for recording and displaying. Typically an electrocardiograph requires an amplification of a couple of thousand times.

Usually recording the electrodiagnostic signals, such as the electrical activity of the heart, is performed in the presence of noise. In case of the electrocardiographic (ECG) signal, two principal sources of noise can be distinguished: the 'technical' caused by the physical parameters of the recording equipment and the 'physiological' representing the bioelectrical activity of living cells not belonging to the area of diagnostic interest (also called background activity). Both sources produce noise of random occurrence, overlapping the ECG signal in both time and frequency domains [1].

This paper proposes a new weighted averaging method, which incorporates empirical Bayesian inference and the expectation-maximization technique. The new method is based on methods proposed in [3] and uses Cauchy distribution. These methods require assumption that p is a positive integer. In numerical experiments performed by the authors, the best results were usually obtained for the parameter $p = 1$ and increasing values of p did not improve performance of the method. That was motivation to extension of the algorithm for some values of $p < 1$. The presented new method is an extension of the methods for $p = 1/2$. In this case the conditional distribution of the signal with respect to some hyperparameter λ is the Cauchy distribution and the unknown parameter λ can be estimated using sample interquartile range.

The paper is divided into four sections. Section 2 describes various methods of signal averaging, from the traditional arithmetic averaging, through the weighted averaging method EBWA [3], which will be treated as the reference method in numerical experiments, to proposed new weighted averaging method. Performance of the new method is experimentally compared with the traditional averaging by using arithmetic mean and also with other empirical Bayesian weighted averaging methods in section 3. Conclusions are given in section 4.

2 Signal Averaging Methods

Let us assume that in each signal cycle $y_i(j)$ is the sum of a deterministic (useful) signal $x(j)$, which is the same in all cycles, and a random noise $n_i(j)$ with zero mean and variance for the i th cycle equal to σ_i^2 . Thus,

$$y_i(j) = x(j) + n_i(j), \quad (1)$$

where i is the cycle index $i \in \{1, 2, \dots, M\}$, and the j is the sample index in the single cycle $j \in \{1, 2, \dots, N\}$ (all cycles have the same length N). The weighted average is given by

$$v(j) = \sum_{i=1}^M w_i y_i(j), \quad (2)$$

where w_i is a weight for i th signal cycle ($i \in \{1, 2, \dots, M\}$) and $\mathbf{v} = [v(1), v(2), \dots, v(N)]$ is the averaged signal.

2.1 Traditional Arithmetic Averaging

The traditional ensemble averaging with arithmetic mean as the aggregation operation gives all the weights w_i equal to M^{-1} . If the noise variance is constant for all cycles, then these weights are optimal in the sense of minimizing the mean square error between \mathbf{v} and \mathbf{x} , assuming Gaussian distribution of noise. When the noise has a non-Gaussian distribution, the estimate (2) is not optimal, but it is still the best of all linear estimators of \mathbf{x} [2].

2.2 Weighted Averaging Method EBWA

The idea of the algorithm EBWA (Empirical Bayesian Weighted Averaging) [3] is based on the assumption that a random noise $n_i(j)$ in (1) is zero-mean Gaussian with variance for the i th cycle equal to σ_i^2 and signal $\mathbf{x} = [x(1), x(2), \dots, x(N)]$ has also Gaussian distribution with zero mean and covariance matrix $B = \text{diag}(\eta_1^2, \eta_2^2, \dots, \eta_N^2)$. The zero-mean assumption for the signal expresses no prior knowledge about the real distance from the signal to the isoelectric line.

The values \mathbf{x} which maximize the posterior distribution over signal and the noise variance (calculated from the Bayes rule) are given by

$$x(j) = \frac{\sum_{i=1}^M \alpha_i y_i(j)}{\beta_j + \sum_{i=1}^M \alpha_i} \tag{3}$$

for $j \in \{1, 2, \dots, N\}$, where $\alpha_i = \sigma_i^{-2}$ and $\beta_j = \eta_j^{-2}$. Maximization the same posterior distribution with respect to α_i gives

$$\alpha_i = \frac{N}{\sum_{j=1}^N (y_i(j) - x(j))^2} \tag{4}$$

for $i \in \{1, 2, \dots, M\}$. Since β_j could not be observed, there is used the iterative expectation-maximization technique. Assuming the gamma prior for β_j with scale parameter λ and shape parameter p for all j , conditional expected value of β_j is given by:

$$E(\beta_j | x(j)) = \frac{2p + 1}{(x(j))^2 + 2\lambda}. \tag{5}$$

Assuming that p is a positive integer, the estimate $\hat{\lambda}$ of hyperparameter λ can be calculated by applying empirical method:

$$\hat{\lambda} = \left(\frac{\Gamma(p)(2p - 1)}{(2p - 1)!!} 2^{p - \frac{3}{2}} \frac{\sum_{j=1}^N |x(j)|}{N} \right)^2, \tag{6}$$

where $(2p-1)!! = 1 \cdot 3 \cdot \dots \cdot (2p-1)$. When the hyperparameter λ is calculated based on first absolute sample moment as in formula (6), the method is described as EBWA.1. The hyperparameter λ can be also calculated based on third absolute sample moment

$$\hat{\lambda} = \left(\frac{\Gamma(p)(2p-3)}{(2p-3)!!} 2^{p-\frac{7}{2}} \frac{\sum_{j=1}^N |x(j)|^3}{N} \right)^{\frac{2}{3}}, \tag{7}$$

which leads to algorithm described as EBWA.3. However, it requires assumption that p is greater than 1.

2.3 Proposed Extension of EBWA Method

Method EBWA described above requires assumption that p is a positive integer. In numerical experiments performed by the authors, the best results were usually obtained with the parameter $p = 1$ and increasing values of p did not improve performance of the method. It was motivation to extension of the algorithm for some values of $p < 1$.

Since the probability distribution function $p(x(j)|\lambda)$ can be written in the form:

$$p(x(j)|\lambda) = \frac{\lambda^p (2p-1)!!}{\Gamma(p) (x(j)^2 + 2\lambda)^{p+\frac{1}{2}}}, \tag{8}$$

it can be observed that for $p = \frac{1}{2}$ the function $p(x(j)|\lambda)$ is Cauchy probability distribution function:

$$p(x(j)|\lambda) = \frac{\sqrt{2\lambda}}{\pi (x(j)^2 + 2\lambda)} \tag{9}$$

with the scale parameter equal to $\sqrt{2\lambda}$ and the location parameter equal to 0. Although all absolute moments of the Cauchy distribution are infinite, the first and third quartiles are the linear functions of scale parameter. It follows from the inverse cumulative distribution function of the Cauchy distribution:

$$F^{-1}(t; \mu, \gamma) = \mu + \gamma \tan \left(\pi \left(t - \frac{1}{2} \right) \right), \tag{10}$$

where μ is the location parameter and γ is the scale parameter.

Thus the hyperparameter λ can be estimated by applying empirical method based on sample interquartile range:

$$\hat{\lambda} = \frac{(\widehat{Q3} - \widehat{Q1})^2}{8}. \tag{11}$$

Therefore the proposed new weighted averaging algorithm can be described as follows, where ε is a preset parameter:

1. Initialize $\mathbf{v}^{(0)} \in R^N$ and set iteration index $k = 1$.
2. Calculate the hyperparameter $\lambda^{(k)}$ using (11), next $\beta_j^{(k)}$ using (5) for $j \in \{1, 2, \dots, N\}$ and $\alpha_i^{(k)}$ using (4) for $i \in \{1, 2, \dots, M\}$, assuming $\mathbf{x} = \mathbf{v}^{(k-1)}$.
3. Update the averaged signal for k th iteration $\mathbf{v}^{(k)}$ using (3), $\beta_j^{(k)}$ and $\alpha_i^{(k)}$, assuming $\mathbf{v}^{(k)} = \mathbf{x}$.
4. If $\|\mathbf{v}^{(k)} - \mathbf{v}^{(k-1)}\| > \varepsilon$ then $k \leftarrow k + 1$ and go to 2, else stop.

3 Numerical Experiments

Performance of the new method was experimentally compared with the traditional averaging by using arithmetic mean and empirical Bayesian weighted averaging methods EBWA described above. In all experiments, using weighted averaging, calculations were initialized as the means of disturbed signal cycles and the parameter ε was equal to 10^{-6} . For a computed averaged signal the performance of tested methods was evaluated by the maximal absolute difference between the deterministic component and the averaged signal (MAX). The root mean-square error (RMSE) between the deterministic component and the averaged signal was also computed. All experiments were run in the R environment for R version 2.4.0 (www.r-project.org).

The simulated ECG signal cycles were obtained as the same deterministic component with added independent realizations of random noise. The deterministic component was ANE20000 from database CTS [5]. A series of 120 ECG cycles was generated with the same deterministic component and zero-mean white Gaussian noise with different standard deviations with constant amplitude of noise during each cycle. Figure 1 presents the ANE20000 ECG signal and this signal with Gaussian noise with standard deviation equal to sample standard deviation of the deterministic component.

In consecutive three experiments there were taken different noise amplitude characteristics varying from $0.05s$ to $2s$, where s is the sample standard

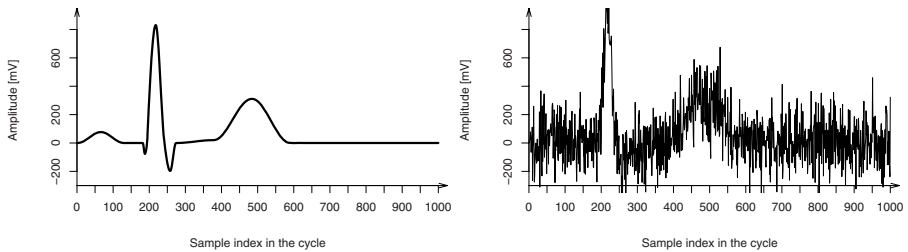


Fig. 1. The example of ECG signal and this signal with Gaussian noise

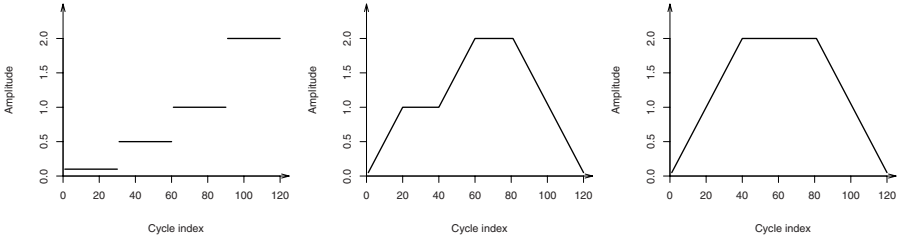


Fig. 2. Noise amplitude characteristics for the first, second and third case

deviation of the deterministic component. Noise amplitude characteristics for the experiments are depicted in Figure 2.

In the first case the amplitude characteristic was step function of cycle index as depicted in Figure 2 (on the left):

$$A(i) = \begin{cases} 0.1s, & i \in \{1, \dots, 30\} \\ 0.5s, & i \in \{31, \dots, 60\} \\ s, & i \in \{61, \dots, 90\} \\ 2s, & i \in \{91, \dots, 120\}. \end{cases} \quad (12)$$

Table 1 presents the RMSE and the absolute maximal value (MAX) of residual noise for all tested methods such as traditional Arithmetic Averaging (AA), empirical Bayesian weighted averaging methods: EBWA.1 and EBWA.3 with parameter p varying from 1 to 3. The table also contains the results for proposed new method of Empirical Bayesian Weighted Averaging using Cauchy distribution (EBWA.C). For all methods, except of traditional Arithmetic Averaging, the number of iterations (NI) is also presented.

In the second case the amplitude characteristic was piecewise linear function of cycle index as depicted in Figure 2 (in the middle):

$$A(i) = \begin{cases} 0.05is, & i \in \{1, \dots, 20\} \\ s, & i \in \{21, \dots, 40\} \\ (0.05(i - 40) + 1)s, & i \in \{41, \dots, 60\} \\ 2s, & i \in \{61, \dots, 80\} \\ (2.05 - 0.05(i - 80))s, & i \in \{81, \dots, 120\} \end{cases} \quad (13)$$

and Table 2 presents results for this case.

In the third case the amplitude characteristic was another piecewise linear function of cycle index as depicted in Figure 2 (on the right):

$$A(i) = \begin{cases} 0.05is, & i \in \{1, \dots, 40\} \\ 2s, & i \in \{41, \dots, 80\} \\ (2.05 - 0.05(i - 80))s, & i \in \{81, \dots, 120\} \end{cases} \quad (14)$$

and Table 3 presents results for this case.

Table 1. Results for the first case

Method	RMSE	MAX	NI
AA	15.45650	57.50675	–
EBWA.C	2.509856	8.927813	7
EBWA.1 (p=1)	2.528694	9.022351	7
EBWA.1 (p=2)	2.533610	9.042611	7
EBWA.1 (p=3)	2.534527	9.045733	7
EBWA.3 (p=2)	2.535590	9.052356	7
EBWA.3 (p=3)	2.536005	9.054720	7

Table 2. Results for the second case

Method	RMSE	MAX	NI
AA	18.15370	56.02270	–
EBWA.C	5.77179	19.67679	24
EBWA.1 (p=1)	5.941526	20.179262	25
EBWA.1 (p=2)	5.977385	20.402368	25
EBWA.1 (p=3)	5.981387	20.439554	25
EBWA.3 (p=2)	5.999454	20.513765	26
EBWA.3 (p=3)	6.005478	20.542843	26

Table 3. Results for the third case

Method	RMSE	MAX	NI
AA	20.59289	72.02635	–
EBWA.C	6.167813	21.472778	28
EBWA.1 (p=1)	6.371921	22.293662	30
EBWA.1 (p=2)	6.423091	22.539492	31
EBWA.1 (p=3)	6.433121	22.582760	31
EBWA.3 (p=2)	6.450463	22.660405	31
EBWA.3 (p=3)	6.458534	22.691401	31

As can be seen in all cases the best results were obtained by using the new EBWA.C method both with respect to the root mean-square error (RMSE) and with respect to the maximal absolute difference between the deterministic component and the averaged signal (RMSE).

4 Conclusion

In this work the new method of weighted averaging of biomedical signal was presented along with the application to averaging ECG signals. Presented new method, which incorporates empirical Bayesian inference, resolves the weighted signal averaging problem and it is an extension of the method proposed in [3] by introducing Cauchy distribution and the unknown parameter is estimated using interquartile range. Application of this method to electrocardiographic (ECG) signal averaging leads to results competitive with other existing methods.

Acknowledgement. This research was supported by the Polish Ministry of Science and Higher Education as the research project N N518 1200 33.

References

1. Augustyniak, P.: Time-frequency modelling and discrimination of noise in the electrocardiogram. *Physiological Measurement* 24, 1–15 (2003)
2. Leski, J.: Application of time domain signal averaging and Kalman filtering for ECG noise reduction. Ph.D. Thesis, Silesian University of Technology, Gliwice (1989)
3. Momot, A., Momot, M., Leski, J.: Bayesian and empirical Bayesian approach to weighted averaging of ECG signal. *Bulletin of the Polish Academy of Science. Technical Science* 55(4), 341–350 (2007)
4. Schamroth, L.: *An introduction to electrocardiography*. Oxford Press, Oxford (1986)
5. International Electrotechnical Commission Standard 60601-3-2 (1999)

An Approach to Estimation of the Angular Eye-Ball Speed Based on the EOG Signal

Tomasz Przybyła, Tomasz Pander, Robert Czabański, and Norbert Henzel

Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland

{Tomasz.Przybyla,Tomasz.Pander,Robert.Czabanski,Norbert.Henzel}@polsl.pl

Summary. In this paper we presented an approach to segmentation of an electrooculography (EOG) signal. For segmentation we have used the elements of the fuzzy set theory. Results obtained in our numerical experiments show usefulness of proposed approach. Our method can be also used for the generating of a learning set necessary for the neural networks or the fuzzy–neural systems training.

1 Introduction

The EOG signal is based on electrical measurements of the potential difference between the cornea and the retina. The cornea–retinal potential creates an electrical field in the front of a head. This field changes in orientation as the eyeballs rotate. The electrical changes can be detected by electrodes which are placed near eyes. It is possible to obtain independent measurements from each of eye. For a healthy man, the movement of eyes is coupled in the vertical direction. Then, it is adequate to measure the vertical motion of single eye together with the horizontal motion of a pair of eyes. The amplitude of EOG signal varies from 50 to 3500 μV with a frequency range of about DC-100 Hz. Eye's behavior is practically linear for gaze angles of $\pm 30^\circ$ [1]. It should be pointed out any biopotential that is measured in the human body is almost always recorded with a noise and often have a non-stationary features. Its magnitude varies in time, even when all possible variables are under control. This means that the variability of the EOG signals depend on many factors that are difficult to determine [1].

The EOG signal can be recorded in a horizontal and a vertical direction of eye movement. This requires six electrodes which are placed in the front of a human face. An eyelid movement (blink) introduces a change in the potential distribution around the eye [2]. Another way to record eye movement signals and eye blinks is to observe different reflections of the emitted infrared light from eyelid and eyeball [3, 2].

The EOG signal contains two different parts. The first part corresponds to the changes of the observed scene. The changes correspond to the rapid variations of the EOG amplitude. The second part of the electrooculography signal corresponds to the saccadic eye movement. Saccadic parts of the EOG signal distinguish on the signal as the small but rather fast changes of amplitude. The

saccade is the fastest movement of an external part of the human body. The peak angular speed of the eye during a saccade reaches up to 1000 degrees per second. Saccades last from about 20 to 200 milliseconds.

Our aim is to detect these parts of the EOG signal, which correspond to the forced change of the observed scene. The main problem is the extraction of these parts from the analyzed signal which contain the highest slopes and correspond to the forced change of the observed scene. The desired parts of the analyzed EOG signal can be found as these parts that are placed between the saccadic parts. This paper focuses on the segmentation problem. As the one of the possible applications, the estimation of angular speed of the eye-ball is presented.

The methods of signal segmentation require the number of segments as an initial parameter [4, 5]. In our approach to the segmentation problem, the number of segments is not required as an initial parameter.

The paper is divided into the following sections: Section 2 describes proposed segmentation and the estimation methods. Section 3 presents obtained results from our numerical experiments. Finally, in section 4 we draw conclusions.

2 Estimation of the Eye-Ball Angular Speed

The EOG signal represents an electric activity of the eyeball muscles. An example of the electrooculography signal has been presented in figure 1. Looking at the example, it can be found that the rapid changes of amplitude correspond to consciously eye move to another part of the scene. Between rapid amplitude changes small and rather fast changes of amplitude can be observed. These parts of the analyzed signal are called saccadic eye motions. Let us define the gradient of the EOG signal as

$$dx(n) = |x(n+1) - x(n)|, \quad (1)$$

where $x(n)$ is the n -th sample of the EOG signal, $1 \leq n \leq N-1$, and N is the length of the signal. Analyzing the gradient signal $dx(n)$ the moment can be found, when *large* amplitudes occur. It matches the change of observed object or part of the scene. Hence, when the gradient values are *small* then we deal with saccadic movements. An example of the EOG gradient signal has been presented in figure 2.

The notions *large* amplitude and *small* amplitude can be interpreted from the fuzzy set theory point of view. So, we can state that the interesting parts of analyzed signal occur when the amplitudes of gradient signal are small.

Many fuzzy clustering methods can be utilized for the estimation of the membership values. Namely, as clustering results we obtain the membership grades for the clustered data, and the cluster prototypes. Unfortunately, after clustering process only the membership grades for the samples from the input data set are known. When the input data set would be changed, for some samples from the data set we could not estimate the membership values. In the worst case, only for few samples from the input data set we could estimate the membership

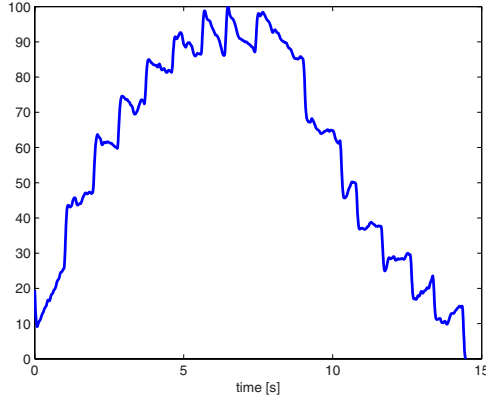


Fig. 1. An example of the EOG signal. The amplitude of the signal has been normalized.

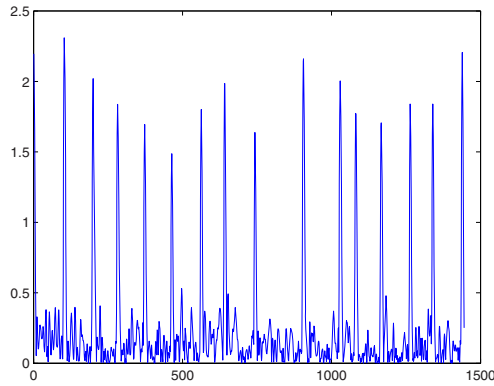


Fig. 2. The gradient of EOG signal

values. So, in this paper, instead of clustering data for each analyzed signal, the $\mathcal{Z}(x)$ membership function has been proposed in the following form:

$$\mathcal{Z}(x) = \begin{cases} 1 & \text{if } x < a, \\ 1 - 2 \left(\frac{x-a}{b-a} \right)^2 & \text{if } a \leq x \leq \frac{a+b}{2}, \\ 2 \left(\frac{b-x}{b-a} \right)^2 & \text{if } \frac{a+b}{2} \leq x \leq b, \\ 0 & \text{if } x > b \end{cases} \quad (2)$$

An example shape of the \mathcal{Z} membership function has been presented in figure 3.

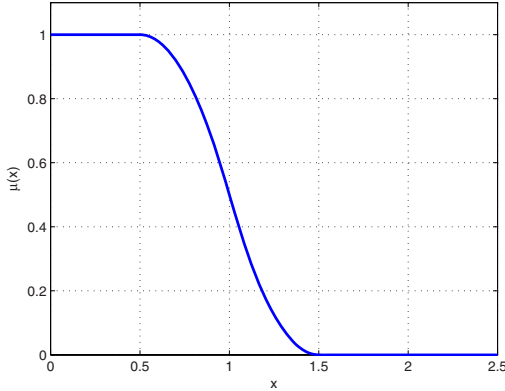


Fig. 3. Shape of the \mathcal{Z} membership function, where $a = 0.5$ and $b = 1.5$

The two parameters a and b can be computed in the following way: for the clustered data set \mathcal{X} and for the sets defined as:

$$\begin{aligned} \mathcal{I}_1 &= \{i : 1 - \mu(x_i) < \varepsilon, 1 \leq i \leq N\} \\ \mathcal{I}_0 &= \{i : \mu(x_i) < \varepsilon, 1 \leq i \leq N\} \end{aligned} \quad , \tag{3}$$

the parameter values can be estimated from

$$\begin{aligned} a &= \frac{1}{|\mathcal{I}_1|} \sum_{i \in \mathcal{I}_1} x_i, \\ b &= \frac{1}{|\mathcal{I}_0|} \sum_{i \in \mathcal{I}_0} x_i, \end{aligned} \tag{4}$$

The $\mu(x_i)$ denotes the membership grade of the i -th sample from the data set obtained in the clustering process, $x_i \in \mathcal{X}$ and $1 \leq i \leq N$, the parameter ε describes the tolerance limit explained further in this section. The notations $|\mathcal{I}_1|$ and $|\mathcal{I}_0|$ correspond the cardinal numbers of \mathcal{I}_1 and \mathcal{I}_0 sets, respectively.

After the clustering process, the membership values are equal to one, only for these samples from the input data set that are equal to the cluster prototype (i.e. the distance between the samples and the prototype is equal to zero) [6, 7]. Hence, for the estimation of the a parameter value, the obtained cluster prototype value corresponds to *small* cluster can be applied. For the methods that utilize the kernel functions a problem occurs for the cluster prototype value determination. Therefore, to avoid such kind of problems, we have proposed another way for estimation the a parameter value.

The proposed equations (3) and (4) can be interpreted as a mean of these samples, that have the membership grades not smaller than $(1 - \varepsilon)$ for the problem of the a parameter value estimation. The membership grades are not higher than the threshold ε during the b parameter value estimation.

The segments (the saccade parts of the EOG signal) are determined as these parts of the analyzed signal, for which the membership values of the *small* fuzzy set of the gradient signal exceed the threshold 0.5. Generally, when the membership grade values exceed $1/c$, where the c is the number of clusters.

After threshold procedure described above, the start point and the end point of the obtained segments are different than expected. So, as the final stage of the processing the EOG signal, the nearest local extrema of the EOG signal have been chosen to the obtained points.

The eye-ball speed can be calculated as:

$$\gamma = \frac{\Delta \Phi}{\Delta t}, \quad (5)$$

where: the $\Delta \Phi$ is the angle range and the Δt is the duration of the eye-ball movement. In our investigation, the $\Delta \Phi$ parameter has been fixed on 10 degrees. The Δt parameter can be computed as the duration of the obtained segment.

Generally, the proposed approach could be described as follows:

- Normalize the EOG signal, increase the amplitude of the signal by 100 times,
- Compute the gradient signal,
- Cluster the amplitudes of the gradient signal into two groups: *small* amplitudes and *large* amplitudes,
- For the *small* amplitudes fuzzy set compute the values of the a and the b parameters of the \mathcal{Z} membership function,
- Choose these samples of the gradient signal which have the membership grade of the *small* amplitudes fuzzy set higher than 0.5,
- Find the local extrema of the EOG signal and chose these, that are nearest to the obtained points in the previous stage,
- Finally, evaluate the angular speed based on (5).

3 Numerical Experiment

In this section, the example demonstrates the performance of the proposed approach. In our investigations, the real, *in vivo* acquired signal have been analyzed. As the data acquisition unit we used the Biopack hardware. For the registration of eye movements we have constructed a segment of LEDs placed on sphere in -40 up to $+40$ degrees with 10 degrees step in horizontal direction. The LEDs were displayed sequentially. In our experiments the OEG signal obtained from three different persons ($P_1 - P_3$) have been used. The proposed method has been applied to the raw signals. All analyzed signals have not been preprocessed. The gradient signal has been multiplied by 100 due to the round-off problems.

As the first stage of the EOG signal segmentation, the gradient signal based on (1) has been estimated. Next, the gradient values have been clustered into two groups corresponding to the *small* amplitude and the *large* amplitude fuzzy sets. As the clustering method, we used the familiar fuzzy c-means clustering

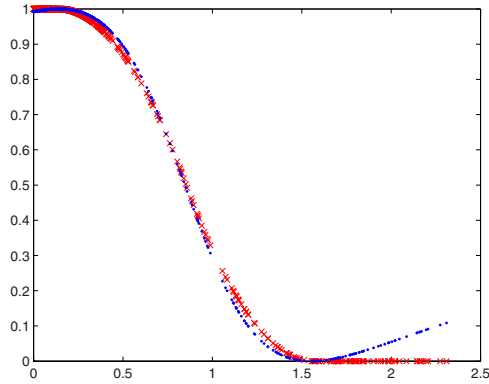


Fig. 4. The comparison between the Z membership drawn by crosses and obtained from clustering process drawn by dots

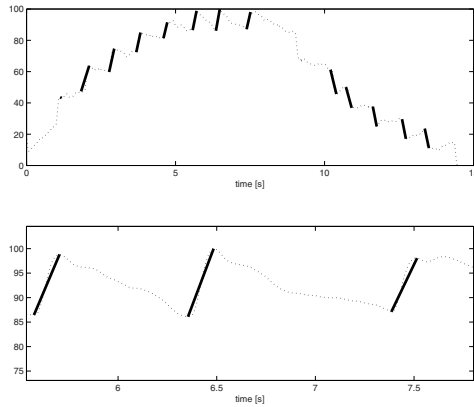


Fig. 5. The segmented EOG signal. The whole signal (a) and the short part of the signal (b). The EOG signal is plotted as the dotted line, and the segments as the solid line.

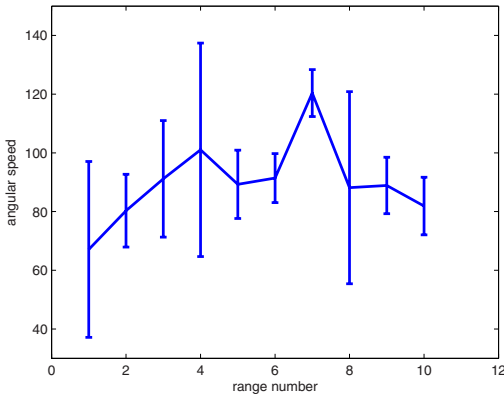
method (FCM) proposed by Bezdek. The following parameters have been fixed for the clustering method:

- the number of clusters $c = 2$,
- the fuzzyfier $m = 2$,
- the tolerance for the FCM method $\epsilon = 10^{-5}$.

After the clustering process, the a and the b parameter values have been estimated. Value of the ϵ parameter has been fixed to $\epsilon = 10^{-3}$. The comparison of the estimated (Z) membership function and the obtained membership values for the *small* data set, have been shown in figure 4.

Table 1. The angular speed of eye-ball for different angular range

No	Angle range	P_1 [$^{\circ}/s$]	P_2 [$^{\circ}/s$]	P_3 [$^{\circ}/s$]
1	$-40^{\circ} - -30^{\circ}$	62.500	90.909	83.333
2	$-30^{\circ} - -20^{\circ}$	71.429	90.909	83.333
3	$-20^{\circ} - -10^{\circ}$	83.333	100.000	125.000
4	$-10^{\circ} - 0^{\circ}$	125.000	111.111	142.857
5	$0^{\circ} - 10^{\circ}$	83.333	90.909	100.000
6	$10^{\circ} - 20^{\circ}$	76.923	90.909	111.111
7	$20^{\circ} - 30^{\circ}$	71.429	83.333	90.909
8	$30^{\circ} - 40^{\circ}$	66.667	76.923	90.909

**Fig. 6.** The average and the standard deviation values of the eye-ball speed for different angle ranges

The obtained segments of proposed algorithm have been shown in fig 5.

Finally, the table 1 shows the obtained eye-ball angular speed. The average and the standard deviation values of the obtained results also in figure 6 have been plotted.

4 Conclusions

In this work, a proposition of an approach to the segmentation of an EOG signal has been introduced. The proposed approach is based on the fuzzy logic. Our aim was to estimate the eye-ball angular speed. The proposed algorithm deals with *human-like* rules, ie. the criterion can be described by linguistic variables such as *small*, or *large*. It makes the algorithm more general.

It should be mentioned, that preprocessing has not been applied. The recorded EOG signals were corrupted by noise, and often the corruption signals have had the non-stationary features.

In our future investigations we concern the preprocessing stage. Specifically, the question of kinds of filters (if any) used for improvement of the final results. Moreover, we touch on the problem of estimation the signal slopes in the saccade parts, and the problem of the influence of blinking should also be considered.

References

1. Barea, R., Boquete, L., Mazo, M., Lpoez, E., Bergasa, L.M.: E.O.G. guidance of wheelchair using neural networks. In: Proc. 15th Int. Conf. on Pattern Recognition, Barcelona (2000)
2. Skotte, J.H., Nøjgaard, J.K., Jørgensen, L.V., Christensen, K.B., Sjøgaard, G.: Eye blinking frequency during different computer tasks quantified by electrooculography. *Eur. J. Appl. Physiol.* 99, 113–119 (2007)
3. Caffier, P.P., Erdmann, U., Ullsperger, P.: Experimental evaluation of eye-blink parameters as a drowsiness measure. *Eur. J. Appl. Physiol.* 89, 319–325 (2003)
4. Moghaddamjoo, A.: Constraint Optimum Well-Log Signal Segmentation. *IEEE Trans. Geoscience and Remote Sensing* 27(5), 633–641 (1989)
5. Moghaddamjoo, A.: Automatic Segmentation and Classification of Ionic-Channel Signals. *IEEE Trans. Biomed. Eng.* 38(2), 149–155 (1991)
6. Bezdek, J.C.: *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum, New York (1981)
7. Pedrycz, W.: *Knowledge-Based Clustering*. Wiley-Interscience, Chichester (2005)

Ensuring the Real Time Signal Transmission Using GSM/Internet Technology for Remote Fetal Monitoring

Krzysztof Horoba, Janusz Wrobel, Dawid Roj, Tomasz Kupka,
Adam Matonia, and Janusz Jezewski

Institute of Medical Technology and Equipment, Biomedical Informatics Dept.,
Roosevelta St. 118, Zabrze, PL 41-800
krzysztof.horoba@itam.zabrze.pl

Summary. This paper presents some aspects of a on-line remote fetal monitoring system based on the GPRS packet data transmission service and WAN network. The system enables analysis, archiving and presentation of signals directly from homes of high-risk pregnancy patients. Since the transmitted signals have to be presented on-line in the surveillance center, possible problems in transmission channel were considered in details. Data stream continuity for visualization is ensured by implementation of incoming data buffer and procedures of its data read-period control.

1 Introduction

Cardiotocography (CTG) is one of the most widely used methods of fetal monitoring, which enables evaluation of the fetal condition during pregnancy and in labour. It relies on the analysis of characteristic fetal heart rate (FHR) patterns in relation to the uterine contractions activity and fetal movements. The goal of electronic fetal monitoring is the assessment of uteroplacental physiology to indicate the adequacy of fetal oxygenation. In traditional cardiotocography the signals recorded and processed by the bedside monitor are presented as printed waveforms. Their visual evaluation is subjective and considerably depends on the experience and knowledge of a clinician. The use of computerized systems for fetal monitoring assures the required objectivity and reproducibility of CTG signal interpretation. Automated analysis allows for more accurate evaluation of the recorded trace due to the computation of instantaneous fetal heart rate variability and classification [8] of the correlation level between the FHR and uterine activity patterns. Centralized fetal monitoring system facilities simultaneous monitoring of many patients hospitalized in the maternity ward as well as in the labour and delivery wards [7]. However, currently there is no possibility to monitor the patient outside the medical centre. The input device is a fetal monitor communicating with the computer via a serial interface. Recorded signals from all fetal monitors, along with results of analysis, are simultaneously presented on the monitoring station.

The use of telemedicine has been increasing worldwide in the last few years. Among the wide range of innovative technology solutions for healthcare, the

telemonitoring is one of the most significant. Telemonitoring becomes common for patients with diabetes or hypertension. Chronic illnesses such as cardiopulmonary disease, asthma, and heart failure or the level of activity of elderly people [9] have also been monitored at a distance. In case of high-risk and post-term pregnancies the cyclic monitoring sessions for follow-up the fetal development process should be carried out. So far continuous medical care requires a hospitalization of pregnant woman even when her health is not at direct risk. It results in high cost of longer hospital stay and psychological discomfort for a patient. The optimal solution seems to be the fetal monitoring at home [5], but in coordination with patient's care personnel from the central surveillance station located in hospital. The general idea of telemonitoring incorporates two concepts of data transmission. The first is the off-line monitoring, when the acquired data are stored in the local memory of remote instrumentation. After the monitoring session is finished, the measurement data are delivered to the surveillance center for analysis. Increasing accessibility to teletransmission technology (classical telephony, GSM, internet) has allowed for development of the on-line monitoring systems [1], where the continuous error-free transmission is assured and the monitoring session is carried out exactly in the same way as in hospital. Nowadays the most flexible data transmission service is the wireless packet transmission (GPRS) offered by GSM network operators [4, 10].

2 Methodology

The overview of the proposed fetal telemonitoring system is presented in Fig. 1. Its two main components are: the Mobile Instrumentations (MI) used for acquisition and transmission of vital signals from fetus, and the Surveillance Center (SC) which is responsible for collecting data, the on-line analysis and archiving of the signals incoming from MIs. Wireless communication is based on the GPRS packet data transmission service provided by GSM operator, and WAN network for data transfer between the GSM and the Surveillance Center. The most obvious way of fetal telemonitoring is to provide the patient with her own MI for the whole time period when the monitoring is recommended [3]. Unfortunately, in this approach each fetal monitor can be used by one patient only. High cost of the fetal monitor and long time of its usage by the patient (with high-risk pregnancy) is a limitation. Reducing the costs, a member of the patient's care team with a fetal monitor and PDA computer could visit patients appointed to be monitored according to a fixed schedule. However, some logistic problems with visit scheduling, especially for large medical centers incorporating numerous patients, should be solved.

2.1 Instrumentation

Single Mobile Instrumentation comprises a standard fetal monitor and Personal Digital Assistant (PDA) as a computer with built-in GSM module, assuring the wireless connection through the WAN network [2]. The PDA software (Fig. 2) is

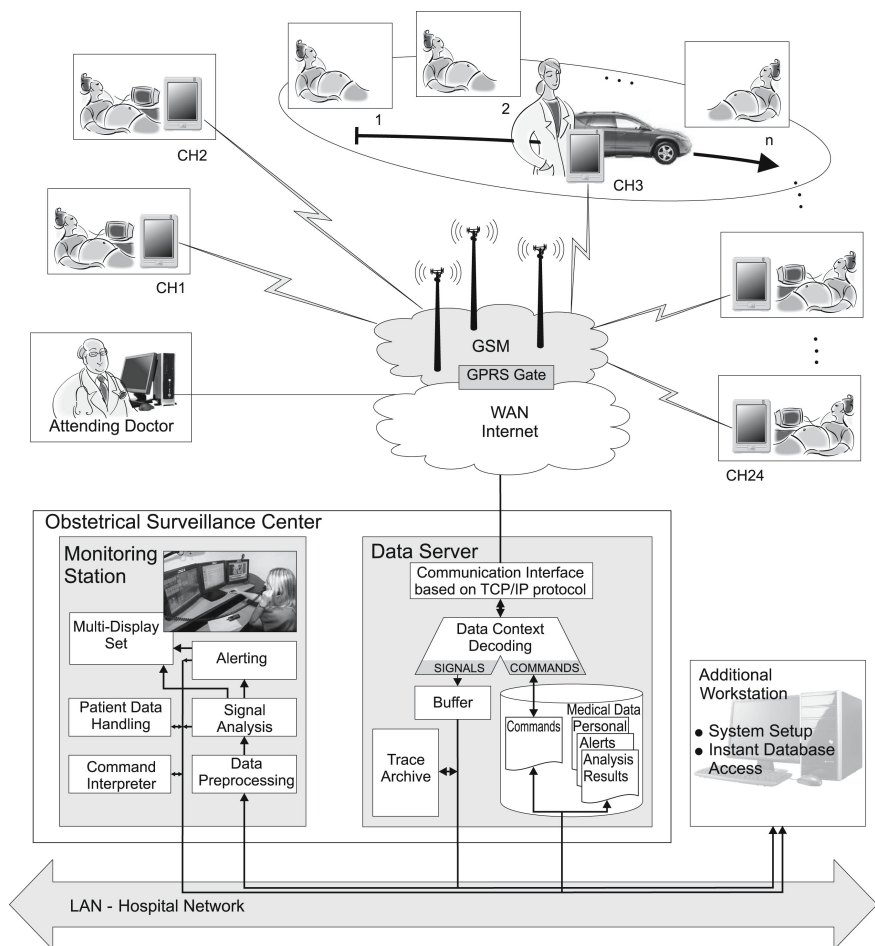


Fig. 1. Block diagram of the essential components that make up the telemedical system for fetal monitoring

responsible for acquisition of measured data from fetal monitor by the RS 232 or USB link, dynamic presentation of the incoming signals and on-line evaluation of their quality. Within the MI various types of fetal monitor can be used, however, due to the required portability of instrumentation, the adapted monitor should be small and light. Since monitor manufacturers prefer their own transmission protocols, the PDA being an interface between fetal monitor and the SC, should make automated identification and adaptation to a given protocol. We distinguished two main protocols: request-answer protocol and automated start-stop one [6]. In the first (e.g. Team monitor, Oxford, G. Britain) the measured signals are send from the monitor output only after receiving the proper request. Despite the type of protocol used, the signals measurements are done every 250 ms.

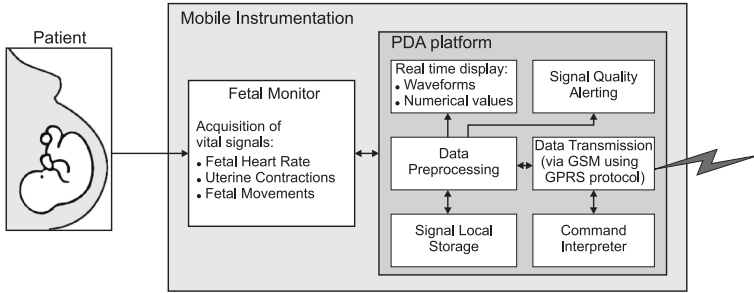


Fig. 2. The structure and interactions between different components of Mobile Instrumentation for remote fetal monitoring

This sampling interval is accepted as a standard. It ensures that no heart beat is lost because 250 ms corresponds to 240 bpm - the upper range of FHR. In a case of the automated start-stop transmission (FM20 monitor, Philips Medical, The Netherlands) the monitor starts sending data frames as soon as it receives a single request for transmission beginning. Each frame contains four samples of registered signals. Surveillance Center contains data server and monitoring station (Fig. 1) communicating each other via hospital LAN. Data server is a dedicated computer with database, the archive of records and the communication module which receives the data from MIs and prepares them for further processing in the monitoring station. The number of displays in the monitoring station depends on the number of MIs predicted to be used. Each display allows for presentation of signals coming from up to eight MIs. Additional workstation, connected to the data server through the LAN, provides an independent and constant access to patients data and acquired signals. Workstation can be used to set up the system, to create paper documentation as well as to process the signals recorded in the off-line mode (eg. in case of total breaking of the communication link). It is possible to access the information stored in the archive from outside the hospital via internet. This feature allows the obstetrician to view the monitoring records at any time he needs. However, due to the personal data protection, the access is permitted only for attending doctor for a given patient.

2.2 Communication between MI and SC

Data exchange between Mobile Instrumentation and Surveillance Center is based on TCP/IP protocol which assures the error-free data transfer. Three types of data blocks have been applied. The command blocks are used for beginning and ending of monitoring, as well as the service of events occurring during transmission. The other types are: blocks with measurement data transmitted by fetal monitor and test blocks without any information content which are sent only to check the connection status. Today, using the GSM technology is relatively inexpensive and makes easy to carry out the monitoring in any place within the

range of the GSM operator's network. However, during the development of our system we met some problems relating to control sharing, ensuring continuous communication between the SC and MI, as well as designing the optimal set of system commands and confirmations. The bottleneck for the overall transmission is the GSM network, where the bandwidth in the worst conditions is limited to 8 kbit/s. However in the presented telemedical fetal monitoring system this value still ensures the correctness of communication because the PDA forms the data stream which does not exceed 1 kbit/s.

Since we assumed that the transmitted signals have to be presented on-line in Surveillance Center, a major issue that has to be considered is the quality of service assured by the GSM operator and internet network, eg. latency, jitter, packet loss rate and bit error rate, on which we have no influence. During the communication the transmission delays can occur due to the overload of base transceiver stations or network infrastructure. Additional time shifts between packets received in SC arise as a result of the internet network rules. When the collection of related packets is routed through the internet, different packets may take different routes, each resulting in a different delay. Moreover, the packet transmission, as opposed to phone connection, does not reserve transmission channel. Data coming from different users are transferred within the given channel. It may take a long time for a packet to reach its destination, because it gets held up in long queues or takes a less direct route to avoid congestion. A packet delay varies with its position in the queues of the routers along the path between source and destination and this position can vary unpredictably (Fig. 3a). Sometimes packets should be retransmitted when the router fails to deliver some of them - the packets arrive when router's buffer is already full. Long-lasting state of router congestion stops the transmission until new route is assigned. Fig. 3b presents temporary stop of the transmission channel during which the mechanism of the TCP protocol is responsible for retransmission of unconfirmed packets. When the stop reason is over, all these packets are transferred at once.

2.3 Buffer Control

The transmission problems have to be corrected in the on-line telemonitoring system to assure continuous data stream for the visualization procedure and signal analysis algorithms. It is necessary to assume the constant time shift between the data sent from the MI and received in the monitoring station. This time shift is implemented in SC using the FIFO register. Buffering eliminates fluctuations in delay of data readings. The buffer size should ensure constant data feeding to the monitoring station (Fig. 4). Big size of the buffer ensures continuity of data flow even in case of long-lasting gaps in transmission. On the other hand, too big size increases the delay between the real signal and the effect of its analysis observed in SC. It is undesirable effect from the point of view of the patient's monitoring. We fulfilled these contradictory requirements implementing the dedicated algorithm for dynamic controlling of the period of data-read from the buffer.

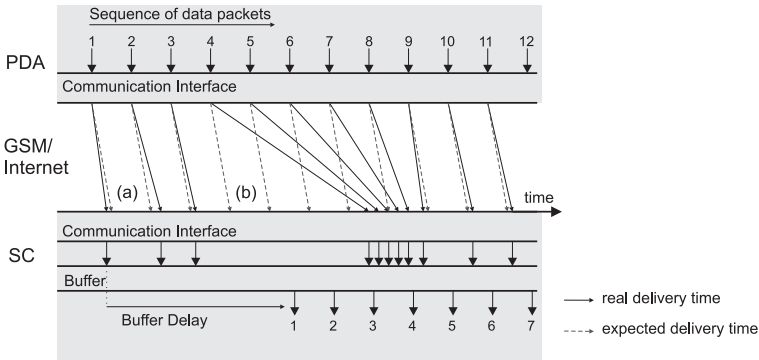


Fig. 3. Transmission of data packets via internet and GSM network with possible disturbances: jitter - difference between real transmission time and its average value (a), and transmission channel blockade leading to packet delivery delay (b)

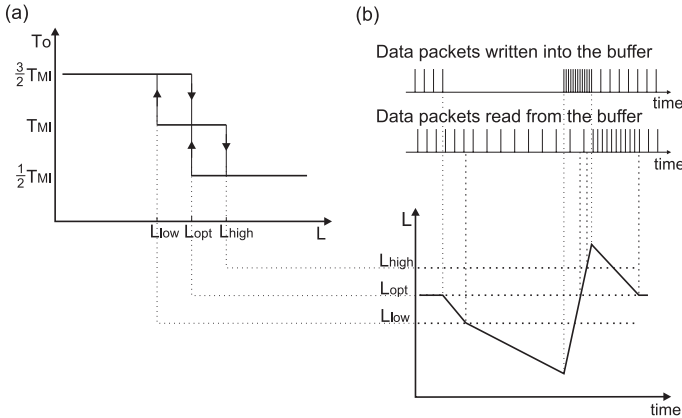


Fig. 4. An illustration of our control algorithm for buffer emptying based on changing the period of data readings from the buffer (a). Example of temporary transmission break during data sending by Mobile Instrumentation (b). Data from MI are transmitted with T_{MI} period, data received from the buffer (for visualisation) with the period T_O . Buffer level indicator - L .

The principles of buffer level control are shown in Fig. 4. Three thresholds are established to control the data-read period. The L_{opt} value describes the total buffer occupancy in normal conditions. The L_{low} and L_{high} values define the data-read period changes. As long as the total buffer occupancy is between L_{low} and L_{high} values, the data-read period is constant and equal to the transmission period of data packets coming from MI. The long-lasting delays in transmission are the reason of reducing the data occupancy in the buffer. If the border value of L_{low} is reached, the data-read period is increased to $T_O = \frac{3}{2}T_{MI}$ which causes

that the speed of CTG signal flowing in monitoring station slightly lowers. In most cases the increasing of data-read period lasts up to a few seconds, and short slowed down fragments are acceptable for clinicians because the speed of traces flowing is rather low (max 3 cm/min). The situation when delayed and current data are simultaneously received results in quick buffer filling. When the buffer occupancy exceeds L_{opt} level the data-read period returns to its initial value T_{MI} . Further rise of occupancy exceeding the L_{high} value causes decrease of data-read period to $T_O = \frac{1}{2}T_{MI}$. Analogically, data packets are read from the buffer with higher rate until the occupancy lowers to the level of L_{opt} .

The buffer control algorithm assures the compromise between buffer size and the time shift between transmitter and receiver. The continuous display of CTG signals may be sustained during the breaks in transmission for even longer than the time shift resulted from the buffer. Another feature of our control mechanism is a lack of waiting time when the buffer is being filled with data. Visualization procedure starts immediately after the first data packets are written into the buffer, but the traces are shifted more slowly to allow the incoming data to fill the buffer. This process continues until the buffer occupancy reaches the L_{opt} level. During the system development we predicted the episodes of data transmission blockade leading to a complete buffer emptying and then to an immediate stop of displaying and analysis of signals. In this case both the receiver and transmitter have to decide if the monitoring should be stopped or one should wait for a possibility of resuming the transmission. If connection can not be resumed during a fixed time, the decision should be made whether to restart monitoring or to switch into the off-line mode. In this mode the acquired data are stored in flash memory and after monitoring completes they can be transmitted to the SC via GSM (when the link is restored) or directly copied into the database using the additional workstation.

3 Conclusion

We hope that telemedical systems for home fetal monitoring based on internet and GSM network will soon become a standard in large medical centers. Technology of wireless packet data transmission assures the remote fetal monitoring everywhere within the range of GSM network. However, the available transmission technologies require development of algorithms to cope with the communication problems relating to dynamic changes of transmission conditions. Continuous data stream for signal visualization is ensured by implementation of incoming data buffer and procedures of its data read-period control. Telemonitoring of pregnant patients at their homes improves access to care, allows to avoid unnecessary hospitalization and makes communication with an attending obstetrician much easier.

Acknowledgement. Scientific work supported by a grant from Iceland, Liechtenstein and Norway through the EEA Financial Mechanism and by financial resources on science in 2007-2010 as a research grant.

References

1. Bai, J., Zhang, Y., Shen, D., et al.: A portable ECG and blood pressure telemonitoring system. *IEEE Eng. Med. Biol.* 18, 63–70 (1999)
2. Choi, J., Yoo, S., Park, H.: A PDA-Based Mobile Clinical Information System. *IEEE Trans.Inf.Technol.Biomed.* 10, 627–635 (2006)
3. Hod, M., Kerner, R.: Telemedicine for antenatal surveillance of high-risk pregnancies with ambulatory and home fetal rate monitoring: an update. *J. Perinat. Med.* 31, 195–200 (2003)
4. Jasemian, Y., Nielsen, L.A.: Design and implementation of a telemedicine system using Bluetooth protocol and GSM/GPRS network, for real time remote patient monitoring. *Technol. Health Care* 13, 199–219 (2005)
5. Jezewski, J., Horoba, K., Gacek, A.: Telemedical system for home monitoring of women with high risk pregnancy. In: *I Int.Conf. on E-Medicine.* (abstract, ,2007)
6. Jezewski, J., Kupka, T., Horoba, K.: Extraction of fetal heart rate signal as time event series from evenly sampled data acquired using Doppler ultrasound technique. *IEEE Trans.Biomed.Eng.* 55, 805–810 (2008)
7. Jezewski, J., Wrobel, J., Horoba, K., et al.: Centralised fetal monitoring system with hardware-based data flow control. In: *Proc. III Conf. MEDSIP*, pp. 51–54 (2006)
8. Jezewski, M., Wrobel, J., Horoba, K., et al.: The prediction of fetal outcome by applying neural network for evaluation of CTG records. In: Kurzynski, M., et al. (eds.) *Advances in Soft Computing Series*, pp. 532–541. Springer, Heidelberg (2007)
9. Lin, B.-S., Chou, N.-K., Chong, F.-C.: RTWPMS: A Real-Time Wireless Physiological Monitoring System. *IEEE Trans.Inf.Technol.Biomed.* 10, 647–656 (2006)
10. Salvador, C.H., Carrasco, M.P., et al.: Airmed-Cardio: A GSM and Internet Services-Based System for Out-of-Hospital Follow-Up of Cardiac Patients. *IEEE Trans.Inf.Technol.Biomed.* 9, 73–85 (2005)

Prediction of Newborn Sex with Neural Networks Approach to Fetal Cardiotocograms Classification

Michał Jezewski¹, Robert Czabanski¹, Krzysztof Horoba²,
Janusz Wrobel², and Janusz Jezewski²

¹ Silesian University of Technology, Institute of Electronics, ul. Akademicka 16,
44-100 Gliwice, Poland
michal.jezewski@polsl.pl

² Institute of Medical Technology and Equipment, Biomedical Informatics
Department, ul. Roosevelta 118, 41-800 Zabrze, Poland

Summary. Cardiotocographic monitoring (CTG) is the primary biophysical method for assessment of the fetal state. It consists in analysis of fetal heart rate variability, uterine contraction activity and fetal movements signal. Visual analysis of printed cardiotocographic traces is difficult so the computerized fetal monitoring systems are a standard in clinical centres. In the proposed work we investigated the ability of the application of artificial neural networks for the prediction of newborn sex using parameters of quantitative description of CTG traces. We examined the influence of input data representation (numerical or categorical) and the influence of the gestational age on the classification quality. We obtained the classification quality at a level of 80% and therefore we may state, that there is rather a strong relation between the fetal gender and the fetal heart rate variability.

1 Introduction

Cardiotocographic monitoring (CTG) is the most common biophysical method used for the assessment of the fetal state. It consists in simultaneous registration and analysis of three signals: fetal heart rate (FHR), uterine contraction activity and fetal movements. Classical visual analysis of printed cardiotocographic traces is rather difficult even for the experienced clinicians. Visual interpretation is highly subjective and may be also characterized by high interobserver and intraobserver disagreement. The FHR variability contains the most important diagnostic information, which is hidden for a naked eye, but can be quantitatively described with a help of dedicated computer-aided systems. Therefore, computerized fetal monitoring systems have become a standard in many clinical centres [4]. But effective methods supporting the diagnostic procedure by automatic evaluation of CTG records are still an important topic of research [5, 6, 8]. In the proposed work we tried to apply neural networks for the prediction of newborn gender (sex) using parameters of quantitative description of antenatal cardiotocographic traces. The results of the other studies on the relation between the fetal gender and its heart rate activity are unequivocal. In [3, 7, 9] no

significant influence of gender on FHR variability was found, whereas in [2] the possibility of such reliance was suggested. In our opinion, the knowledge about such relation is very crucial, because a possible influence of the gender on the FHR variability should be taken into a consideration during the interpretation of CTG signals, as expected to be different for male and female fetuses.

During our research we tried to answer some important questions, which may occur during the application of computational intelligence (like neural networks) as a tool for the vital signals classification. We investigated, how the accuracy of neural network classification can be affected by different representations of the input variables. Additionally, we examined the influence of the gestational age on the gender prediction quality.

2 Research Material and Methodology

The research material was obtained from the computerized fetal surveillance system MONAKO [4]. It included results of quantitative analysis of 749 traces from 103 patients, with 261 (35%) relating to male and 488 (65%) to female newborns. The traces duration was from 21 to 340 minutes (an average 44 minutes). The records were registered between 28th and 42nd week of gestation. As input variables for the neural network we used 17 parameters of quantitative description of the CTG signals. Fifteen parameters describe the fetal heart rate signal in time domain: baseline (MeanFHR, Fluct), acceleration/deceleration patterns (ACCfreq, DECfreq), beat-to-beat variability (STVB, STV, DLM_BB, DLM, STIBB, STI) as well as the long-term variability (LTV, OSC, OSC_0, OSC_III, LTI). Additionally, the frequencies of recognized uterine contractions (CONfreq) and fetal movements (MOVfreq) were used. The detailed description of the input parameters is presented below:

- MeanFHR - mean value of the FHR baseline.
- Fluct - difference between the maximum and minimum value of FHR. This parameter describes fluctuations of the baseline around its mean value.
- ACCfreq - frequency of acceleration patterns [number/hour]. Acceleration is an increase of FHR signal above the baseline lasting at least 15 seconds with amplitude of at least 15 bpm.
- DECfreq - frequency of deceleration patterns [number/hour]. Deceleration occurs when FHR signal decreases below the baseline for at least 15 seconds with amplitude of at least 15 bpm.
- STVB - describes absolute differences between successive T_{R-R} intervals.
- STV - describes absolute differences between T_{R-R} intervals averaged over 2.5 seconds epochs.

For STVB and STV a mean value for each one minute trace fragment is calculated and finally the mean value for a whole trace is determined.

- LTV - describes differences between the maximum and minimum value of T_{R-R} intervals averaged over 2.5 seconds epochs.

- OSC (oscillation amplitude) - describes differences between the maximum and the minimum value of FHR.

For LTV and OSC a mean value (weighted mean for LTV) of differences in each one minute trace fragment is calculated.

- OSC_0 - percentage of oscillation time with the amplitude below 5 bpm in relation to a whole trace time.
- OSC_III - percentage of oscillation time with the amplitude above 25 bpm in relation to a whole trace time.
- DLM_BB - mean value (for a whole trace) of Yeh's DI [10] short-term indices values, calculated for one minute trace fragment.
- DLM - mean value (for a whole trace) of Yeh's DI [10] short-term indices values, calculated for one minute trace fragment, but instead of successive T_{R-R} intervals, the averaged values of T_{R-R} intervals over 2.5 seconds epochs are used.
- STIBB - mean value (for a whole trace) of the inter quartile ranges of angles obtained by putting successive T_{R-R} intervals from one minute trace fragment on the coordinate system [10].
- STI - mean value (for a whole trace) of the inter quartile ranges of angles obtained by putting averaged values of T_{R-R} intervals over 2.5 seconds epochs from one minute trace fragment on the coordinate system [10].
- LTI - mean value (for a whole trace) of the inter quartile ranges of segments obtained by putting averaged values of T_{R-R} intervals over 2.5 seconds epochs from one minute trace fragment on the coordinate system [10].
- CONFreq - frequency of uterine contractions [number/hour].
- MOVfreq - frequency of fetal movements [number/hour].

The neural network output was defined as two-state variable defining the newborn gender: male or female. The gender was obtained from the newborns data forms.

Statistica Neural Networks 7.1 (StatSoft, Inc.) software was used for the developing of artificial neural networks. We used two most common types of networks [1]: Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF). For both of them, we changed the number of neurons in a hidden layer to find a network topology with the highest generalization ability. The number of neurons was changed from 2 to 9 with a step of 1 in case of the MLP networks and from 2 to 50 with a step of 4 for RBF networks. For MLP networks we used sigmoid activation function and for RBF a standard Gaussian radial function. During the learning of the MLP networks we applied 1000 epochs of steepest descent gradient algorithm with constant learning rate equals to 0.01. The centres of the RBF neurons were determined by K-Means algorithm and the radii by K-Nearest Neighbors algorithm. Each network was trained 50 times, with the cases randomly assigned to three subsets: training, validating and testing. The number of training examples was two times greater than the number of examples in validating and testing subsets. The sizes of subsets and the number of traces from both classes (male and female newborns) in each subset were constant in

Table 1. Details of antenatal and intrapartum groups

Group	Week of pregnancy	Newborn sex	
		Male	Female
I	33-36	106	179
II	34-37	80	204
III	35-38	54	225
IV	36-41	47	231
V	Labour	137	149

each trial. In order to evaluate the quality of the classification we used a mean value as well as a standard deviation of the correct assignments in testing subset calculated for all 50 trials.

Two main numerical experiments were performed in the presented work. The first one concerns problems, which may occur during development of neural networks based system for CTG signals evaluation. The second one concerns the classification of CTG traces. In the first experiment, the influence of data set representation on the classification quality was examined. The input variables were converted from numerical to categorical values. The conversion referred to the physiological ranges of CTG parameters provided by the clinical expert [10]. The physiological range for a given input variable was a function of gestational age. The conversion was possible only for eleven parameters of quantitative description of CTG traces. We set the value 0 for parameters within the physiological range and value 1 for parameters outside that range. During the second experiment we examined the influence of gestational age on classification quality, because with the progress of pregnancy, the parameters describing quantitatively acquired signals may change. Four groups of traces registered in similar gestational age (from 33rd to 41st week) and one group of traces registered during the labour were created. The groups of antenatal traces overlapped in order to make their sizes similar to the size of labour group (286 cases). The details of antenatal and intrapartum groups are given in Table 1.

3 Results

The usefulness of the designed neural networks was expressed by the classification quality, which describes the percentage of correctly classified cases in a testing subset. All the statistical parameters were calculated for 50 trials for every network. In case of the MLP networks and the numerical representation of data, an increase of mean values of the classification quality was observed with the increase of the neuron number up to 6. For the number of neurons greater than 6 the quality of classification decreased. However the single increase was noticed for 9 neurons in a hidden layer. During the learning of the RBF networks when using the numerical representation of inputs we observed a trend increasing classification quality (up to 34 neurons) with a one decrease for 22 neurons. For

Table 2. Classification results obtained for different input data representations

	MLP			RBF		
	No. of neurons	Mean (SD)	Min-Max [%]	No. of neurons	Mean (SD)	Min-Max [%]
Num.	6	73.9% (3.3)	66-80	34	70.5% (3.1)	62-78
Categ.	9	66.8% (2.9)	61-73	10	62.8% (3.8)	55-70

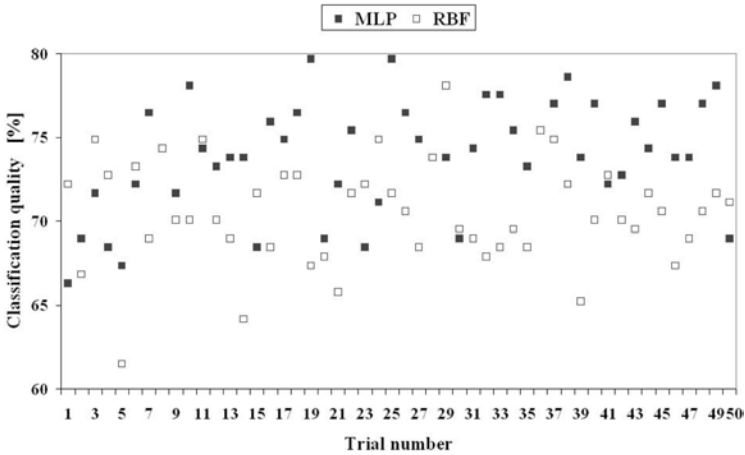


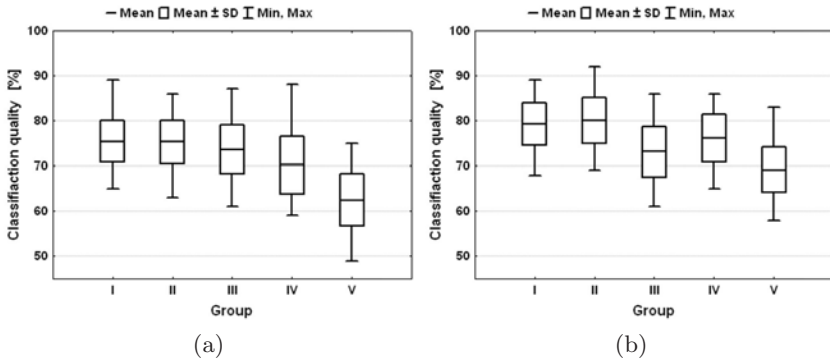
Fig. 1. The classification accuracy for the MLP and the RBF networks with the highest generalization ability obtained for 50 different trials

the MLP networks and the categorical representation of data we observed the best classification quality for 9 neurons in a hidden layer. The best learning results of RBF networks for the categorical data were obtained for 10 neurons. The comparison of the results obtained for two different input data representations (numerical or categorical) and for the networks structures with the highest generalization ability is presented in Table 2. The categorical representation of data led to the decrease (by about 7%) of the classification accuracy. For that reason, we used only numerical representation of data in our further experiments. Also the RBF networks provided a little worse classification quality (by about 4%) in comparison to the MLP networks. Figure 1 presents the prediction quality obtained for the numerical representation of the input data using the MLP and the RBF networks with the highest generalization ability. The level of dispersion is similar for both networks, but it may be also observed, that the MLP network generally leads to the higher classification quality.

In the second group of experiments we investigated the influence of gestational age on classification. The results (for networks structures with the highest

Table 3. Results of gender prediction for traces registered in different weeks of pregnancy and during labour

Group	MLP			RBF		
	No. of neurons	Mean (SD)	Min-Max [%]	No. of neurons	Mean (SD)	Min-Max [%]
I	6	79.3% (4.7)	68-89	34	75.4% (4.6)	65-89
II	6	80.1% (5.1)	69-92	34	75.3% (4.8)	63-86
III	5	73.1% (5.5)	61-86	22	73.6% (5.4)	61-87
IV	6	76.2% (5.3)	65-86	34	70.2% (6.4)	59-88
V	6	69.1% (5.0)	58-83	6	62.5% (5.7)	49-75

**Fig. 2.** Classification quality of the RBF (a) and MLP (b) networks for different gestational age

generalization ability) of the gender prediction for CTG signals recorded in different weeks of pregnancy and during labour are shown in Table 3. The best classification results were obtained for the signals registered in the early weeks (from 33rd to 37th) of pregnancy. Considering the RBF networks we observed the decrease of classification quality with the increase of the gestational age (see Fig. 2(a)). There is no such relation in case of the MLP networks (Fig. 2(b)). Again, the results obtained for the RBF networks were slightly worse than for the MLP networks, with the exception of the pregnancy group III. Standard deviations of classification quality were similar for both types of networks. The best results were obtained for group I and II, for which the classification quality of the MLP networks reached the level of 80%. Definitely, the worst results were observed for signals acquired during labour (group V). It is worth to notice that in the case of the RBF networks, it was enough to use only six neurons in a hidden layer to get the reasonable prediction results in group V.

There are differences (with the exception of group V) between numbers of traces relating to male and female newborns (Table 1), but it is impossible to

make them equal if we want to have similar number of all traces in each group during learning. However, these differences seem not to have an influence on classification quality, because the worst results were obtained not for group IV (the biggest disproportion), but for the intrapartum traces, where trace numbers for male and female newborns were almost equal. In our opinion the relation between the classification results and gestational age comes from the cardiotocographic traces features, not from the data set structure. The results of our experimentation show the decrease of the gender prediction quality with the higher gestational age and suggest the strongest differences in the FHR variability between genders at the early stage of the fetal development.

4 Conclusions

In the presented work we investigated the possibility of the prediction of newborn gender basing on parameters of quantitative description of cardiotocographic signals using neural networks. We tried to find the appropriate data set as well as the network structure. We examined two types of artificial neural networks: MLP and RBF. In both types, number of neurons in a hidden layer was changed to find the network structure ensuring the best generalization ability. Each network was trained 50 times, with random cases assignment to three subsets: training, validating and testing. Such approach increases the reliability of the prediction quality evaluation and eliminates coincidences. The best classification accuracy at level of 80% was obtained for the MLP neural networks.

We investigated also the influence of the input data representation on the classification process. Our experiments show that converting the inputs from numerical to categorical values results in a decrease of the prediction accuracy. We tried to assess the influence of gestational age on classification precision as well. Four groups of antenatal records and one group with signals registered during labour were tested. The highest classification accuracy was obtained for traces recorded in earlier weeks of gestation, the worst prediction of newborn gender was observed for the intrapartum CTG recordings.

We may state, that there is rather a strong relation between the fetal gender and the fetal heart rate variability. The results of our experiments indicate also that the differences in the FHR variability between male and female fetuses at the early stage of the fetal development are the most significant.

Acknowledgement. This work was supported in part by the Ministry of Sciences and Higher Education resources in 2007-2009 under Research Project R1302802.

References

1. Czogala, E., Leski, J.: Fuzzy and Neuro-Fuzzy Intelligent Systems. Physica-Verlag, Berlin (2000)
2. Dawes, N.W., Dawes, G.S., Moulden, M., Redman, C.W.G.: Fetal heart rate patterns in term labour vary with sex, gestational age, epidural analgesia, and fetal weight. American Journal of Obstetrics and Gynaecology 180, 181-187 (1999)

3. Druzin, M.L., Milton, H.J., Edersheim, T.G.: Relationship of baseline fetal heart rate to gestational age and fetal sex. *American Journal of Obstetrics and Gynaecology* 154, 1102–1103 (1986)
4. Jezewski, J., Wrobel, J., Horoba, K., Kupka, T., Matonia, A.: A Centralised fetal monitoring system with hardware-based data flow control. In: *Proc. of III International Conference MEDSIP Glasgow*, pp. 51–54 (2006)
5. Jezewski, M., Wrobel, J., Horoba, K., Gacek, A., Henzel, N., Leski, J.: The prediction of fetal outcome by applying neural network for evaluation of CTG records. In: Kurzynski, M., Puchala, E., et al. (eds.) *Advances in Soft Computing Series*, pp. 533–540. Springer, Heidelberg (2007)
6. Jezewski, M., Wrobel, J., Labaj, P., Leski, J., Henzel, N., Horoba, K., Jezewski, J.: Some practical remarks on neural networks approach to fetal cardiotocograms classification. In: *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5170–5173 (2007)
7. Lange, S., Van Leeuwen, P., Geue, D., Hatzmann, W., Gronemeyer, D.: Influence of gestational age, heart rate, gender and time of day on fetal heart rate variability. *Medical and Biological Engineering and Computing* 43, 481–486 (2005)
8. Magenes, G., Signorini, M.G., Arduini, D.: Classification of cardiotocographic records by neural networks. In: *Proc. of the IEEE International Joint Conference on Neural Networks*, vol. 3, pp. 637–641 (2000)
9. Ogueh, O., Steer, P.: Gender does not affect fetal heart rate variation. *British Journal of Obstetrics and Gynaecology*, 1312–1314 (1998)
10. Sikora, J.: Digital analysis of cardiotocographic traces for clinical fetal outcome prediction. *Clinical Perinatology and Gynaecology Supplement* 21 (in polish, 2001)

Coping with Limitation of Bedside Measurement Instrumentation for Reliable Assessment of Fetal Heart Rate Variability

Janusz Wrobel, Janusz Jezewski, and Krzysztof Horoba

Institute of Medical Technology and Equipment, Biomedical Informatics Dept.,
Roosevelta St. 118, Zabrze, PL 41-800
januszw@itam.zabrze.pl

Summary. Estimation of the instantaneous variability of the fetal heart rate (FHR) is affected by the autocorrelation techniques commonly used in Doppler ultrasound channel of today's fetal monitors. Considerably decrease of short-term variability have been noted, which is quite surprising because the fetal monitors determine the fetal heart rate with quite satisfying accuracy in relation to the reference direct fetal electrocardiography. The aim of this work was to recognize a source of errors and to develop the method for correction of the indices describing the FHR variability for a given type of fetal monitor. The proposed correction relies upon removing of the constant error component, which has been assigned to an averaging nature of the autocorrelation function. Although the remaining random error component is still not too satisfactory considering the instantaneous values, a significant improvement of reliability of the fetal heart rate variability measurement was confirmed in case of a global one-hour trace assessment.

1 Introduction

Evaluation of instantaneous variability of fetal heart rhythm is considered as a valuable predictor of good fetal state. In automated analysis of the fetal heart rate (FHR) signal two main components of instantaneous variability: short- and long-term are evaluated quantitatively using various numerical indices [8]. In time domain analysis they have been defined on a basis of time intervals T_i calculated between consecutive heart beats in fetal electrocardiogram (FECG) [2, 7]. In turn, present-day computerised fetal monitoring systems [6] analyse the FHR signal which is transmitted from bedside monitors equipped with Doppler ultrasound-based technology (US). However, this approach cannot ensure as high accuracy of heart beat detection as the reference fetal electrocardiography. This affects directly the accuracy of calculation of the instantaneous FHR values expressed in beats per minute according to the equation: $FHR_i[bpm] = 60000/T_i[ms]$. Therefore, the algorithms for computation of FHR variability indices are expected to be particularly sensitive to the acquisition method of the fetal heart rate signal.

This problem has been widely discussed, however variety of evaluation and classification methods used makes difficult to estimate definitively an influence

of the signal acquisition method on the reliability of automated analysis of the FHR variability [1]. Error in time positioning of the events corresponding to consecutive fetal heart beats has decisive influence on the accuracy of the T_i intervals measurement. Segments of the Doppler envelope signal, corresponding to successive heart beats are characterised by continuous change of their shape and location of the peaks [9, 11]. Using new generation fetal monitor - with an autocorrelation function - Lawson et al. [10] obtained error of the short-term variability assessment: -35%. A minus sign means that the ultrasound approach has decreased significantly the short-term variability in relation to the direct electrocardiography. We noted earlier [7] that the US method is able to determine T_i interval with the accuracy of about 2.5 ms in relation to the reference electrocardiography, which is 0.7% of the typical measured T_i value of 440 ms. It was stated also that such accuracy is enough for both visual analysis and automated recognition of slow-wave patterns of FHR signal: baseline, acceleration and deceleration. The influence was observed only on evaluation of the instantaneous FHR variability. The long-term variability indices were characterised by the mean error of about -5%, which does not distort considerably the clinical assessment of FHR trace. However the mean error of -22% obtained for the set of short-term variability indices was unacceptable.

Considerable decrease of the short-term variability indices caused by the ultrasound approach actually makes doubtful the use of this technique for clinical assessment of FHR trace, when it should rely on automated analysis of the beat-to-beat variability. There are many bedside monitors of different types still being used in hospitals. Their work is more than satisfying for visual interpretation of the recorded traces. But the computerized monitoring system have become today's standard in fetal surveillance. Since the systems have to cooperate with the input devices being already in use we had to consider a possibility to improve the reliability of the FHR variability analysis which resulted in the proposed correction of the FHR variability indices.

2 Methodology

The research material comprised simultaneously recorded intrapartum FECG and FHR signals from the ultrasound method. The FHR signal of 0.25 bpm resolution was provided by a new generation fetal monitor. Electrocardiogram was captured directly from a fetal head and fed to data acquisition board in laptop computer. Analogue to digital conversion was performed with the sampling frequency of 2 kHz and the resolution of 12 bits. Reference FHR signal was computed from the fetal electrocardiogram in an off-line mode. Two different algorithms were applied to determine T_i as the R-R intervals, and all pairs of values significantly differing were excluded. This limited the error of the reference T_i intervals to 1 ms. Finally, after rejection of the segments with signal loss, 185 minutes of traces were qualified for further analysis.

Analysing the results [7] as well as basing on experiments concerning fetal monitors' operation, the most probable sources of the errors of the FHR

variability quantification have been defined. Algorithms for T_i values calculation applied in today's fetal monitors are based on autocorrelation technique [3]. These algorithms do not detect consecutive heart beats but they only estimate an averaged periodicity of the signal within a window analysed. This process requires the window including at least two T_i intervals. However, to obtain an evident dominating peak of the autocorrelation function from noisy ultrasound signal and to consider the maximum possible beat-to-beat changes of the interval value (e.g. during the deceleration pattern), a wider window comprising three beats has to be used. This leads to the averaging of T_i intervals. However the real number of the averaged intervals is even higher due to a simplified iteration algorithm usually used in bedside monitors for autocorrelation computation.

Short-term FHR variability describes the beat-to-beat changes whereas the long-term one concerns the tendency of these changes in longer time period (usually in one-minute epochs). Considering these definitions an averaging of T_i intervals is supposed to cause a small decrease of values of long-term indices as well as a significant decrease of short-term indices. The variability indices LTI and STI proposed by de Haan's [2] are the most commonly used. Only de Haan's STI index was selected for detailed analysis because, as it was noted previously [7] it had the largest error equal to 39%, while the mean error obtained for the set of short-time variability indices was -22%.

Due to the nature of ΔSTI and its range, this work concerns analysis of the influence of the ultrasound-based measurement method on an estimation of the short-term FHR variability only. This error is proportional to a variability range which is indicated by a slope of the regression line from Fig. 1a. In the previous monitors the errors caused by shape distortion of patterns corresponding to successive heart beats in Doppler envelope were random and resulted mainly from measurement conditions. While in new generation monitors, the error originating in averaging caused by the autocorrelation is connected with a given monitor type with built-in autocorrelation algorithm. As for the random errors from the Doppler envelope distortion they are sufficiently reduced by the autocorrelation algorithm. Having the reference signal as the event series corresponding to the fetal heart beats detected in the fetal electrocardiogram as well as T_i intervals determined using these events, experiment was performed based on modeling the reference signal to simulate the entire error of the new generation monitors and to relate it to the real error of the US method obtained for the research material collected.

The location distortion was simulated for the events corresponding to consecutive heart beats. The values from random generator of the uniform distribution were added to each reference time point defining the fetal heart beat occurrence. The random generator was run with range being changed from $< -1; +1 >$ to $< -10; +10 >$ ms. Than simulation of the error caused by averaging of the intervals in autocorrelation procedure was carried out. Obtained T_i intervals from distorted reference signal were averaged over different numbers of consecutive intervals - from two to eight in each iteration. In order to evaluate a range of distortion in the monitor, the modeling was repeated with various combinations of the levels of distortions and averaging. In every experiment the index

was calculated within one-minute segments of both the reference signal and the modeled one. Dispersion of the absolute index error (the difference between the distorted values and reference) has been shown in relation to the reference signal. Additionally, the regression line as well as mean values of the absolute and relative errors were calculated. The parameters were compared with the parameters obtained for real errors of the STI index from the ultrasound approach.

The obtained results concerning a tendency of STI errors in relation to the degree of the applied distortions made possible to correct the error components appropriate for the US method, and connected with built-in autocorrelation algorithm exclusively. This correction relied on recalculation of the erroneous index values by the use of the linear regression parameters estimated for ΔSTI errors as a function of the reference values STI_{REF} :

$$\left. \begin{array}{l} \Delta STI = a \cdot STI_{REF} + b \\ \Delta STI = STI - STI_{REF} \\ STI^* \rightarrow STI_{REF} \end{array} \right\} \Rightarrow STI^* = \frac{STI}{a + 1} + b \quad (1)$$

where the STI^* is the variability index corrected with linear regression parameters a (slope) and b (intercept) estimated using the reference values.

The remaining error component was estimated after the correction of the STI index. It is obvious that correction of the indices calculated for the same signals acquired with US approach, that have been used to estimate the linear regression parameters, results in zero value of the mean error of US index in relation to the reference. The only component left will be a small random error. Therefore, final verification of the proposed correction method had to be carried out using another database of FHR signals. They were recorded simultaneously using other type of fetal monitor based on Doppler ultrasound technology, and developed system [5] for indirect fetal electrocardiography from maternal abdomen. The signals were split into two separated groups. The correction parameters estimated for one group were used to correct STI indices originally calculated for the other group and vice versa.

3 Results

It has been noted that the values of the absolute errors of STI index obtained for the research material are spread along the regression line (Fig. 1a, Table 1) of the slope equal to -0.542 ($RE = 1.2$). Thus the index calculation error caused by the US method is proportional to the reference value of the index ($r > 0.8$ with $p < 0.02$). This error has a minus sign, which means that the reference value is higher than the value obtained for US, and it increases (as for the absolute value) with the index value increase. The constant component of the error is defined by the intercept of the regression line equal to 0.93 (Fig. 1a).

In the experiment, various jitter distortions as well as averaging were applied. Efforts were made to get dispersion of the STI index error as similar as possible to the results obtained for fetal monitor (Fig. 1b). This was achieved in wide range of heart beat detection distortion from $< -1; +1 >$ ms to $< -10; +10 >$ ms, but

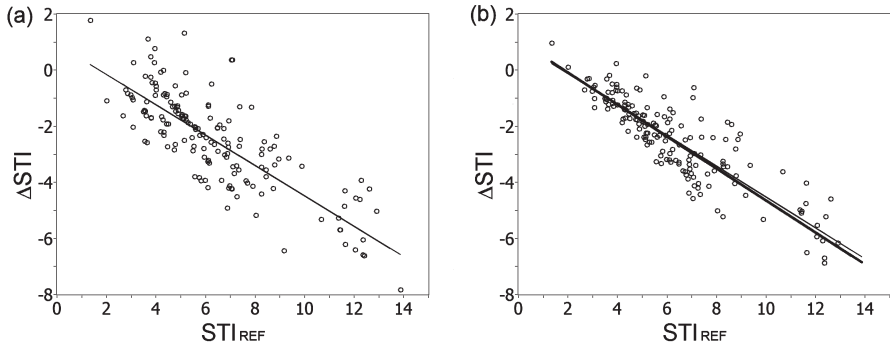


Fig. 1. A scatter plot showing the relation between the absolute error of de Haan's index STI calculated from: (a) US signal, (b) reference signal after adding the jitter of $< -5; +5 >$ and its averaging over five intervals exclusively, against the reference index value. Additional thin lines in (b) represent the regression lines determined for the index errors for US channel (a).

Table 1. Statistical parameters of the STI index and its error determined using the US and after adding to the reference signal distortions: jitter of $< -5; +5 >$ ms and averaging over five intervals (J5A5)

Data sets	STI	Δ STI	δSTI [%]	Linear regression		
				Intercept	Slope	RE^a
J5A5	3.85 ± 1.49^b	2.50 ± 1.71	-39.4	0.93	-0.547	1.3
US	3.88 ± 1.57	-2.47 ± 1.76	-38.9	0.93	-0.542	1.1

^a Residual Root Mean Square Error, known as standard deviation of data about the regression line, ^b Mean \pm SD.

for only the one averaging period - performed within five-interval window. Statistical parameters obtained are almost identical with the real ones (Table 1). The above experiment showed also, that the main source of the errors of the short-term variability indices determination is the averaging effect being characteristic for the autocorrelation procedures. Since it is strictly connected with a given type of fetal monitor, the STI index may undergo to correction for the ultrasound channel method associated with the particular autocorrelation algorithm. Such correction can be carried out using the parameters of the regression line determined experimentally using the scatter plot of relationship between the absolute determination error of a given variability index and the reference value of the index (1).

The last stage was aimed at verification of the efficiency of the correction performed. Values of STI index were recalculated according to the regression line parameters (Fig. 2). After cancellation of the constant error component, the dispersion of the differences around the zero level appeared to be uniform

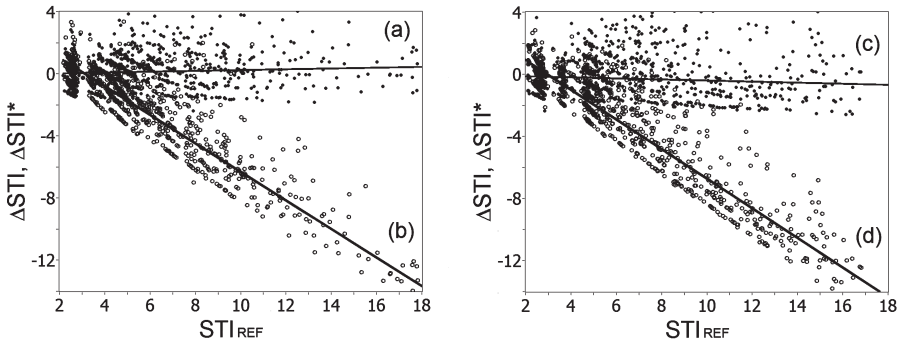


Fig. 2. Results obtained from the validation procedure. Scatter plots show the short-term variability index error ΔSTI^* for the signal group I (a) after correction using the regression line parameters determined for the errors ΔSTI for the group II (d) and vice versa (c)(b).

Table 2. Statistical parameters of the corrected STI index and its errors, where: the indices for real ultrasound signals (US) were corrected using the regression parameters from the same signals (US*), the indices calculated for the first verification group of signals (US I) were corrected using the regression parameters from the second group (US* I/II) and vice versa - the indices from the second group (US II) were corrected using the first group (US* II/I)

Data sets	STI	ΔSTI	δSTI [%]	Linear regression		
				Intercept	Slope	RE^a
US*	6.39 ± 2.80^b	0.00 ± 1.03	0.0	0.00	0.000	1.03
US	3.88 ± 1.57	-2.47 ± 1.76	-38.9	0.93	-0.542	1.10
US* I/II	6.09 ± 3.40	0.46 ± 1.43	8.2	0.00	0.048	1.43
US I	3.54 ± 1.47	-2.09 ± 2.77	-37.1	2.77	-0.899	1.43
US* II/I	6.11 ± 3.60	-0.11 ± 1.44	-1.7	0.00	-0.047	1.44
US II	3.31 ± 1.46	-2.91 ± 3.41	-46.8	2.78	-0.948	1.42

^a Residual Root Mean Square Error, known as standard deviation of data about the regression line, ^b Mean \pm SD.

(Table 2) and not to have evident trend, which indicated a pure random nature of the error. The proposed method for correction of the results of determination process of the FHR variability indices was verified basing on two additional independent groups of signals. The correction parameters determined for the one group were used to correct indices calculated for the second group and vice versa. In both cases (Table 2) the mean error for the STI index corrected was not equal to zero but it was decreased considerably: for the first arrangement from -37.1% to +8.2%, and for the second one from -46.8% to -1.7% (Fig. 2). There was no relationship between the error of the index and its reference value and the error dispersion was uniform (SD = 1.44).

4 Conclusions

Using the Doppler ultrasound method to acquire the fetal heart rate signal, the main problem is the detection of consecutive heart beats in time point of their true occurrence. In modern fetal monitors an application of autocorrelation function for signal processing considerably limited influence of unstable shape of the Doppler envelope and its dynamic changes on the accuracy of T_i intervals determination. Measurement error has decreased to 2.5 ms [7], which is quite satisfactory for visual trace interpretation. This accuracy is also enough for computerised analysis but still only with exclusion of short-term variability. It was shown, that the index error depends mainly on averaging window width, which was followed by the change of regression line slope calculated between the index error and the index value. In turn, the influence of the random error - heart beats location in time - goes to minimum. It may be assumed, that the regression line parameters determined experimentally for a given monitor type are independent from measurement conditions, and they can be used to correct the index value for any other signal recorded by monitor of this particular type.

The proposed method for correction of the indices describing quantitatively the short-term fetal heart rate variability is our next step to increase the reliability of the FHR signal acquired from fetal monitors based on the Doppler ultrasound technology with built-in autocorrelation procedure [4]. This correction relies upon cancellation of the constant error component, which is a result of an averaging nature of the autocorrelation function. This is reflected by the mean value close to zero (Fig. 2). This ensures the improvement of global evaluation of FHR variability for the entire patient's monitoring session, whose duration is usually 60 minutes, because the final variability is computed as a mean value over particular one-minute variability values. The remaining random component of the averaging process reached value $SD = 1.44$, which corresponds to 24% of typical value of $STI = 6.0$ (Table 2). This is not too satisfactory result considering the instantaneous values of the index, particularly if these values are close to the alerting thresholds. Nevertheless, in case of global one-hour evaluation of the fetal heart rate variability a significant improvement of its reliability has been achieved.

Acknowledgement. This work was supported in part by the Ministry of Sciences and Higher Education resources in 2007-2009 under Research Project R1302802.

References

1. Dawes, G.S., Visser, G.H.A., Goodman, J.D.S., Redman, C.W.G.: Numerical analysis of the human fetal heart rate: The quality of ultrasound records. *Am. J. Obstet. Gynecol.* 141, 43-52 (1981)
2. De Haan, J., van Bommel, J.H., Versteeg, B., et al.: Quantitative evaluation of fetal heart rate patterns I. Processing methods. *Europ. J. Obstet. Gynecol.* 3, 95-102 (1971)

3. Jezewski, J., Horoba, K., Wrobel, J., et al.: Monitoring of mechanical and electrical activity of fetal heart. Determination of the FHR. *Arch. Perinat. Med.* 8, 33–39 (2002)
4. Jezewski, J., Kupka, T., Horoba, K.: Extraction of Fetal Heart Rate Signal as Time Event Series from Evenly Sampled Data Acquired Using Doppler Ultrasound Technique. *IEEE. Trans. Biomed. Eng.* 55, 805–810 (2008)
5. Jezewski, J., Matonia, A., Kupka, T., Wrobel, J.: Fetal monitoring with online processing of electrocardiographic signals. *IFMBE Proc 3rd Eur. Med. Biol. Eng. EMBEC 11*, 1–5 (2005)
6. Jezewski, J., Wrobel, J., Horoba, K., et al.: Centralised fetal monitoring system with hardware-based data flow control. In: *3rd IET Int. Conf. MEDSIP 2006*, pp. 1–4 (2006)
7. Jezewski, J., Wrobel, J., Horoba, K.: Comparison of Doppler ultrasound and direct electrocardiography acquisition techniques for quantification of Fetal Heart Rate variability. *IEEE. Trans. Biomed. Eng.* 53, 855–864 (2006)
8. Kubo, T., Inaba, J., Shigemitsu, S., Akatsuka, T.: Fetal heart variability indices and the accuracy of variability measurements. *Am. J. Perinat.* 4, 179–186 (1987)
9. Kupka, T., Jezewski, J., Matonia, A., et al.: Timing events in Doppler ultrasound signal of fetal heart activity. In: *Proc 26th IEEE EMBS Int. Conf. 2004*, pp. 337–340 (2004)
10. Lawson, G.W., Belcher, R., Dawes, G.S., Redman, C.W.G.: A comparison of ultrasound (with autocorrelation) and direct electrocardiogram fetal heart rate detector systems. *Am. J. Obstet. Gynecol.* 147, 721–722 (1983)
11. Shakespeare, S.A., Crowe, J.A., Hayes-Gill, B.R., et al.: The information content of Doppler ultrasound signals from the fetal heart. *Med. Biol. Eng. Comput.* 39, 619–626 (2001)

Relationships between Isopotential Areas in EEG Maps before, during and after the Seizure Activity

Hanna Goszczyńska¹, Marek Doros¹, Leszek Kowalczyk¹, Krystyna Kolebska¹, Stanisław Dec², Ewa Zalewska¹, and Jan Miszczak²

¹ Institute of Biocybernetics and Biomedical Engineering PAS, 4, Trojdena str., 02-109 Warsaw

hania.goszczyńska@ibib.waw.pl

² Military Institute of Aviation Medicine, 54, Krasińskiego str., 01-955 Warsaw

Summary. The aim of the study was to analyse relationships between isopotential areas in sequences of EEG top maps recorded before, during and after seizure activity. Method of the analysis of isopotential areas in sequences of EEG maps consists on the comparison of the changes of values of areas before and after seizure activity and estimation of mutual changes of the areas for the extreme isopotential regions during seizure activity. Results of the study performed on two groups of totally 17 subjects suggest that selected image features like area of given range of potentials and the analysis of the relationships between areas of the isopotential regions before, during and after seizure episode reveal the differences in considered groups of subjects.

1 Introduction

Brain Electrical Activity Mapping BEAM is a routine method used in electroencephalographic (EEG) examinations for visualization of values of the different parameters characterizing the bioelectrical brain activity. The mapping procedures are accessible in the majority commercial devices for the routine EEG examinations, and they are commonly applied in clinical practice [1, 2, 3, 4, 5]. These maps are automatically compared with the reference maps for the control groups of patients. However, the methods of maps comparison available in the specialized commercial electroencephalographic systems are still not sufficient for demanded expectations of medical doctors in the diagnostic usefulness.

The aim of the study was to develop the methods and criteria for the evaluation of variability of the brain bioelectric activity allowing both the improvement of the visual evaluation of the EEG maps, and their quantitative analysis and comparison.

2 Material and Method

Methodology of the examinations comprises the elaboration of the set of the features describing the geometrical properties and mutual localization (topologicalization) of the respective regions in the images (EEG maps) representing the

distribution of the electrical potential on the head surface measured in a given instants of a time. Applying that features and examination of the selected similarity criteria of those regions were carried out. In particular, the analysis of the change of the areas of isopotential regions (A_n , where n indicates the symbol of range of potentials) and the determining the relationships between A_n in sequences of EEG maps before, during and after the seizure activity were taken into consideration.

The present approach concerns the analysis of the relationships between A_n for the whole maps in the sequence of the *top maps*.

The data necessary to the investigations were acquired using the system NeuroScan 4.3. Examinations have been carried out for 17 subjects divided into two groups (common numbering for both groups). The first group (I) was consisted of the 10 clinically healthy subjects with the seizure activity. The second group (II) comprised the seven patients with epilepsy [6, 7]. The start point and the end point of the seizure activity episode in EEG records were determined by clinician.

The material used in the examinations comprised the sequence of 1000 amplitude maps taken from the 10 s EEG record before, during and after the seizure activity episode for each subject (Fig. 1). The dimensions of images of those maps were 628×790 pixels in 17 colors scale referred to the 17 ranges of values of electrical potentials in scale from $-20\mu V$ to $20\mu V$. The maps were generated in instances of 10 ms each.

Fig. 1 shows the example of images from the sequence of 1000 maps referring to 10 s EEG recorded signal chosen from three periods:

- images 150 - 179 before seizure activity episode (frame numbers: 1 - 400),
- images 450 - 479 during seizure activity episode (frame numbers: 401 - 549),
- images 850 - 879 after seizure activity episode (frame numbers: 550 - 1000).

Present approach concerns the analysis of the changes and the relationships between areas A_n for the ranges of minimum and maximum potentials denoted as A_{-20} and A_{20} , respectively. The analysis of changes of A_n in the sequence of the maps was performed using the normalized histograms of images of the *top maps*. Fig. 2 shows the normalized histogram of the map no 150 shown in Fig. 1.

The values of areas for all ranges of potentials were calculated on a basis of the normalized histograms for all maps in the given sequence. Fig. 3a shows the normalized values of areas A_n for all ranges of potentials (17 ranges) for sequence of maps from number 1 to 1000 that partially were presented in Fig. 1. Fig. 3b shows values of areas A_{-20} and A_{20} for the same sequence.

Present method of the analysis of the relationship between A_n in sequences of EEG maps consists on the comparison of the changes of values of areas taken before and after seizure activity and estimation of mutual changes of A_n for the ranges of minimum and maximum potentials during seizure activity.

So, the procedure was performed in two steps:

1. calculation of the difference between the average areas A_{-20} and A_{20} before and after the seizure activity episode,
2. calculation of the values of the mutual ratio of areas A_{-20} to A_{20} during the seizure activity episode.

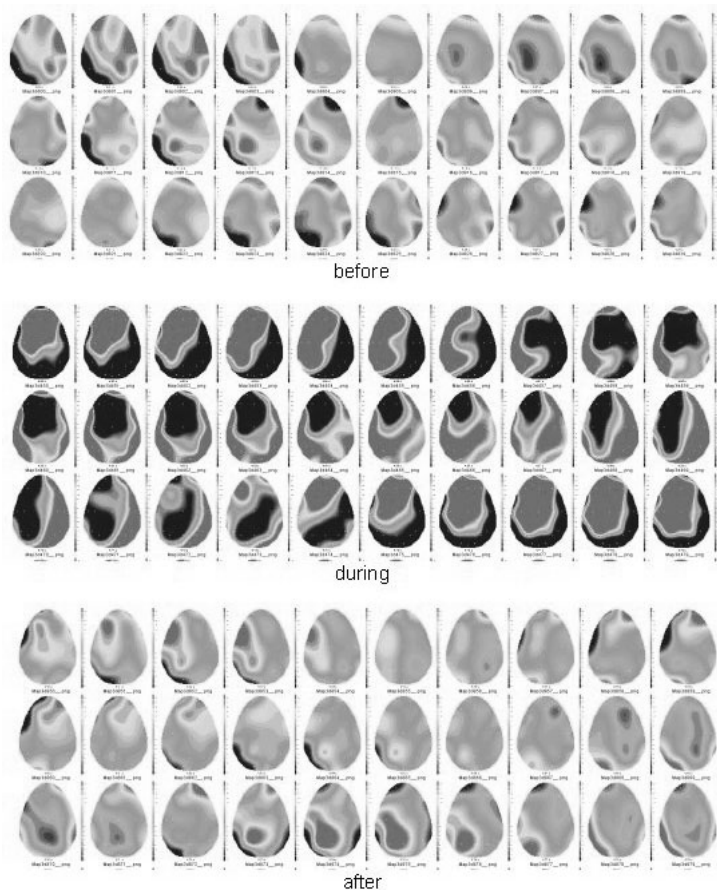


Fig. 1. Selected images from EEG maps sequence before (frame numbers: 150 - 179), during (frame numbers: 450 - 479) and after (frame numbers: 850 - 879) seizure activity episode for patient no 3 from II group

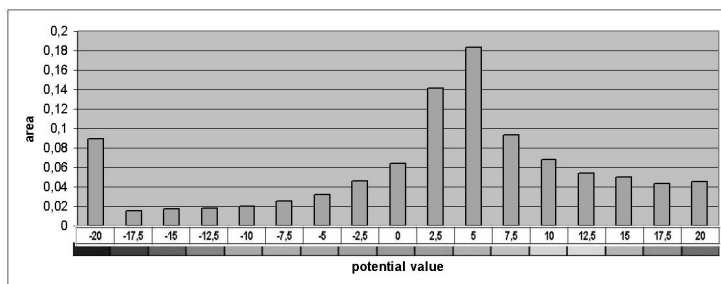


Fig. 2. Normalized histogram of the map no 150 shown in Fig. 1

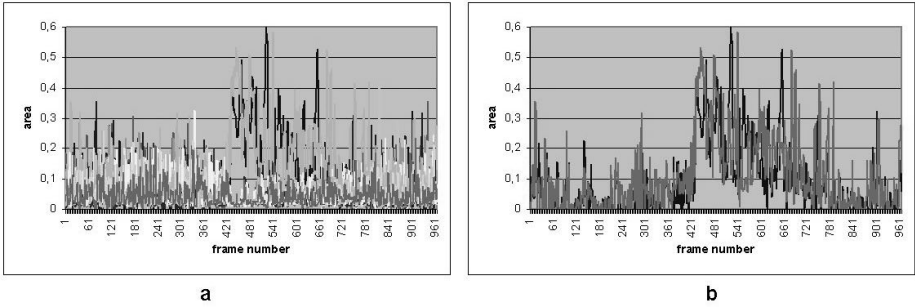


Fig. 3. The normalized values of areas A_n for all ranges of potentials (a) and for the areas A_{-20} and A_{20} (b)

2.1 Relationship between the Areas A_{-20} and A_{20} before and after the Seizure Activity Episode

Fig. 4a shows average values of the areas for all ranges of the potentials before and after the seizure activity episode for values of areas A_n presented in Fig. 3a and the modules of differences between average values of the areas A_n before and after the seizure activity (Fig. 4b).

Then the values of sum of modules of the differences of average areas A_{-20} and A_{20} were calculated.

2.2 Relationship between the Areas A_{-20} and A_{20} during the Seizure Activity

Estimation of the relationship between isopotential areas during the seizure episode was based on the analysis of the values of mutual ratios of the areas A_{-20} and A_{20} . Fig. 5a shows the example of the values of areas A_{-20} and A_{20} during the seizure activity.

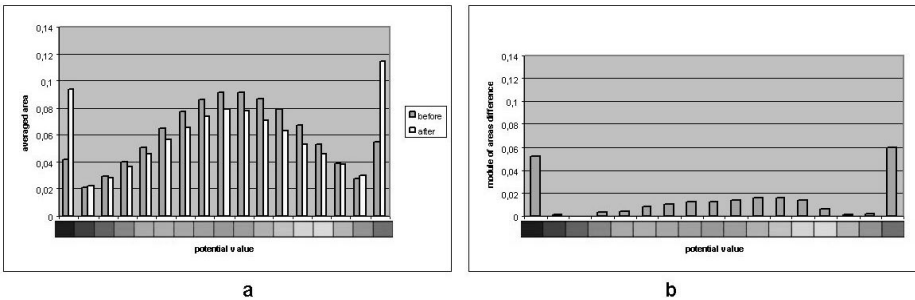


Fig. 4. Average values of the areas A_n for all ranges of potentials before and after the seizure activity episode for values of A_n presented in Fig. 3 (a) and modules of differences of average areas A_n (b)

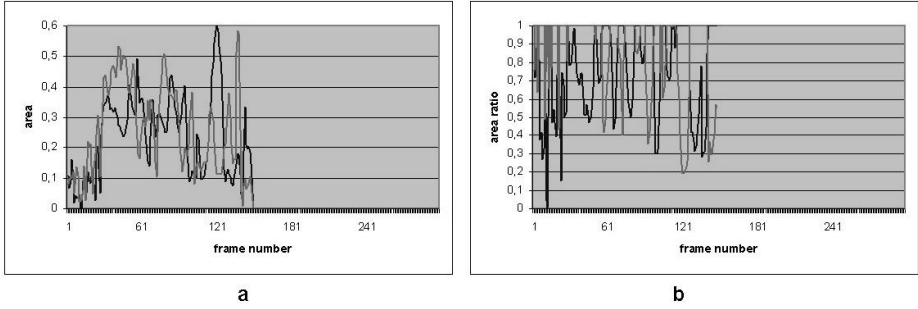


Fig. 5. The normalized values of areas A_{-20} and A_{20} during seizure activity episode (part of curves presented in Fig. 3b) for frame numbers: 401 - 549 (a), the values of ratios for the range from 0 to 1

Two ratios were calculated:

- ratio of the area values regarding $-20\mu V$ potential range to the area value regarding $20\mu V$ potential range $R_{br} = A_{-20}/A_{20}$,
- ratio of the area value regarding $20\mu V$ potential range to the area value regarding $-20\mu V$ potential range $R_{rb} = A_{20}/A_{-20}$.

There were analysed the values of ratios for the range from 0 to 1 (Fig. 5b).

3 Results

Preliminary results of analysis of relationships between the areas A_{-20} and A_{20} before and after the seizure activity episode for 17 subjects divided into two groups were reported earlier in [8, 9]. In point 3.1 the most important results of the examinations carried out are presented in details. In point 3.2 the values of ratio of the areas A_{-20} and A_{20} during the seizure activity episodes for the example of healthy subjects and epilepsy patients are shown.

3.1 Relationship between the Areas A_{-20} and A_{20} before and after the Seizure Activity Episode

Fig. 6 shows the examples of normalized values of the areas A_{-20} and A_{20} before and after the seizure activity for both groups of subjects.

For the healthy subjects the values of the areas A_{-20} and A_{20} before and after seizure activity do not present the differences whereas for patients the values of the areas after seizure activity are higher.

Fig. 7 presents the values of sums of modules of differences of the average areas A_{-20} and A_{20} before and after the seizure activity episode for both groups of subjects.

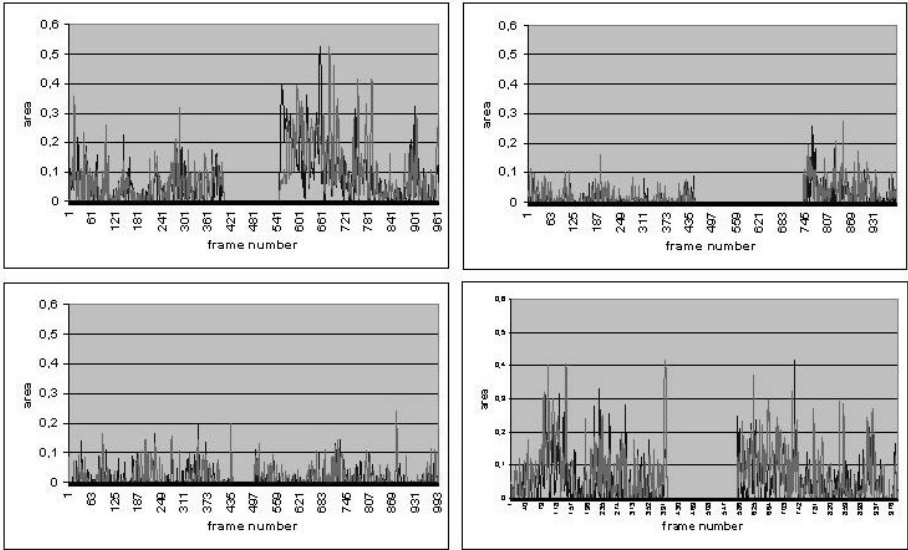


Fig. 6. Examples of the normalized values of the areas A_{-20} and A_{20} before and after the seizure activity episode for group II (patients: no 3 - upper left, no 4 - upper right) and for group I (subjects: no 8 - bottom left, no 16 - bottom right) [7, 8]

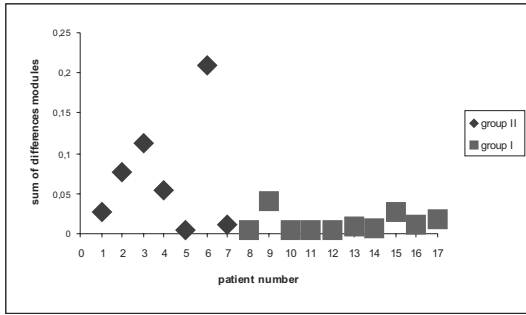


Fig. 7. Sums of modules of differences of the average areas A_{-20} and A_{20} before and after the seizure activity episode for both groups of subjects [8, 9]

3.2 Relationship between the Areas A_{-20} and A_{20} during the Seizure Activity

Fig. 8 shows examples of the normalized values of the areas A_{-20} and A_{20} during the seizure activity episode for the same subjects as in Fig. 6.

Examples of the values of ratio of the areas A_{-20} and A_{20} during the seizure activity episodes for the healthy subjects and epilepsy patients are shown in Fig. 8. For each subjects two ratios were calculated:

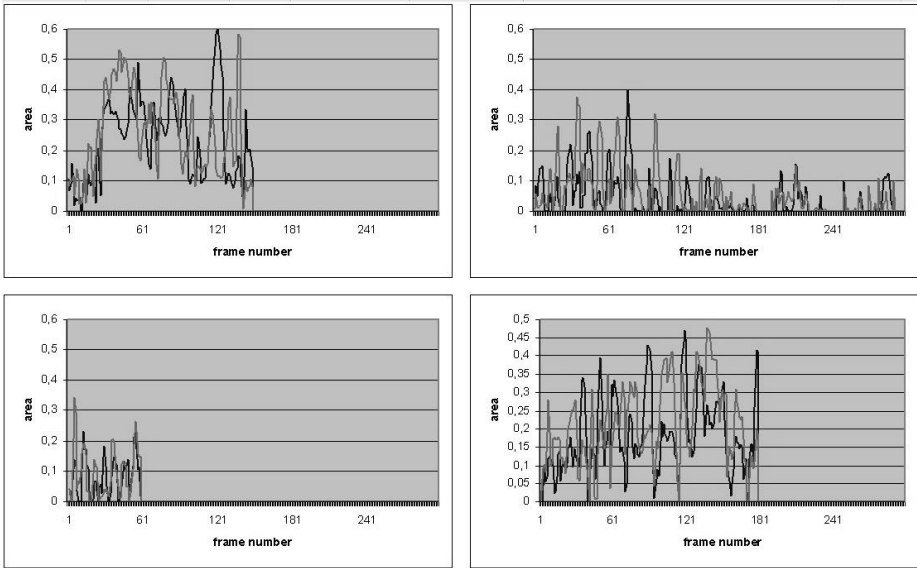


Fig. 8. Examples of the normalized values of the areas A_{-20} and A_{20} during the seizure activity episode for group II (subjects: no 3 - upper left, no 4 - upper right) and for group I (patients: no 8 - bottom left, no 16 - bottom right)

- ratio R_{br} (area values regarding $-20\mu V$ potential range to area value regarding $20\mu V$ potential range) (Fig. 9 left diagrams),
- ratio R_{rb} (area value regarding $20\mu V$ potential range to area value regarding $-20\mu V$ potential range) (Fig. 9 right diagrams).

For the healthy subjects the values of ratio of the areas A_{-20} and A_{20} during seizure activity are more variable than for the patients. For the patients the shapes of the envelope curves of the values of the ratio seems to be more regular.

4 Discussion

Results of the study performed on the groups of totally 17 subjects suggest that selected image features like area of given range of potentials and the analysis of the relationships between the areas A_{-20} and A_{20} for whole maps in the sequence of *top maps* before, during and after seizure episode may be a foundation for differentiation of EEG maps.

However, it should be noticed that the important issue in this analysis was setting the amplitude scale for EEG maps of subjects. A range from $-20\mu V$ to $20\mu V$ was chosen which seemed to be the most convenient. From the other side, it may be the possible source of errors in the results of analysis of the A_n . In what follows, two examples are discussed.

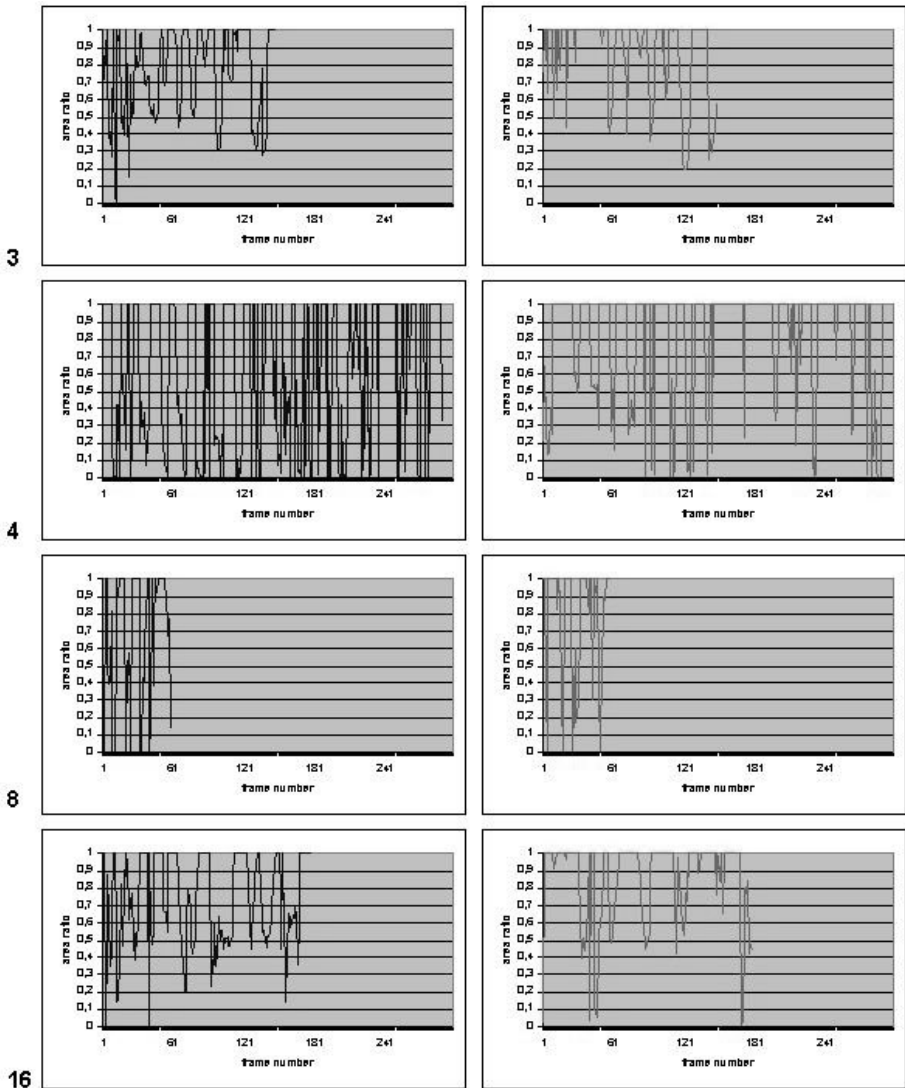


Fig. 9. Example of the values of ratios R_{br} (left diagrams) and R_{rb} (right diagrams) for group II (patients: no 3, no 4) and for group I (subjects: no 8, no 16)

For the patient no 5 from group II (see Fig. 7) in the majority of images of maps before and after seizure activity episode there are no regions regarding $-20\mu V$ or $20\mu V$ ranges of potentials and the normalized values of the areas for the ranges of minimum and maximum potentials before and after the seizure activity episode have very low amplitude (Fig. 10). For this case the scale should be narrower.

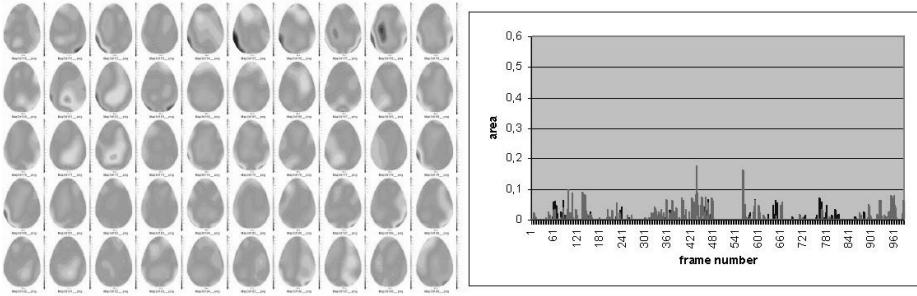


Fig. 10. Selected maps (frame numbers: 150 - 199) from the period before seizure activity (frame numbers: 1 - 480) for patient no 5 from group II and the normalized values of the areas A_{-20} and A_{20} before and after the seizure activity episode (see Fig. 6)

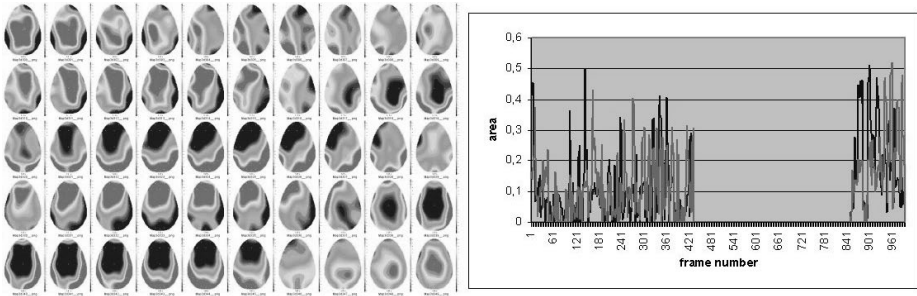


Fig. 11. Selected maps (frame numbers: 350 - 399) from the period before seizure activity (frame numbers: 1 - 430) for patient no 6 from group II and the normalized values of the areas A_{-20} and A_{20} before and after the seizure activity episode (see Fig. 6)

For the patient no 6 from group II (see Fig. 7) the majority of images of maps before and after seizure activity episode are similar to the maps characteristic for seizure activity period and the normalized values of the areas for the ranges of minimum and maximum potentials before and after the seizure activity episode demonstrate high amplitudes (Fig. 11). For this case the scale should be wider.

Analysis of the normalized values of the areas A_{-20} and A_{20} during the seizure activity episode (Fig. 8) does not indicated the differences between both groups of subjects. Analysis of the values of the ratio of the areas A_{-20} and A_{20} during the seizure activity episode (Fig. 9), suggests, however, that the differences may be characterized by the shape of the envelope curves of the ratio values.

References

1. Morihisa, J.M., Duffy, F.H., Wyatt, R.J.: Brain electrical activity mapping (BEAM) in schizophrenic patients. *Arch. Gen. Psychiatry* 40, 719–728 (1983)
2. Lehman, D.: From mapping to the analysis and interpretation of EEG/EP maps. In: Maurer, K. (ed.) *Topographic Brain Mapping of Eeg and evoked potentials*, pp. 53–75 (1989)
3. Li, L., Yao, D.: A new method of spatio-temporal topographic Mapping by correlation coefficient of k-means cluster. *Brain Topography* 19(4), 161–176 (2007)
4. Miszczak, J.: Neurofizjologiczne implikacje badań stanów czynnościowych mózgu metodą mappingu potencjałów wywołanych. *Polski Przegląd Medycyny Lotniczej* 3, 333–349 (2003)
5. Dec, S., Miszczak, J., Zalewska, E.: Obraz QEEG klinicznie zdrowych kobiet i mężczyzn w wieku 18-30 lat-kandydatów do lotnictwa. *Polski Przegląd Medycyny Lotniczej* 10, 209–226 (2004)
6. Kowalczyk, L., Dec, S., Zalewska, E., Miszczak, J.: Progress on a Study of Spontaneous EEG Activity in Records of Epilepsy Patients with Interictal Discharges and Epileptiform Discharge. In: *Healthy Subjects, Proc. of 7th European Congress on Epileptology, Helsinki* (2006)
7. Kowalczyk, L.: Badanie dynamiki aktywności bioelektrycznej mózgu w warunkach stymulacji z zastosowaniem metod oceny podobieństwa sygnałów, *Rozprawa doktorska, IBIB PAN, Warszawa* (2006)
8. Goszczyńska, H., Kowalczyk, L., Świdorski, B., Kolebska, K., Dec, S., Zalewska, E., Miszczak, J.: Nowe kryteria analizy ilościowo-porównawczej map EEG. In: *Mat. Konf. V Symposium Naukowe: TPO 2006, Serock*, pp. 152–157 (2006)
9. Goszczyńska, H., Kowalczyk, L., Świdorski, B., Kolebska, K., Dec, S., Zalewska, E., Miszczak, J.: Ocena ilościowo-porównawcza sekwencji czasowych map EEG w celu ustalenia kierunków propagacji pobudzenia. In: *XV Krajowa Konferencja Biocybernetyki i Inżynierii Biomedycznej, Wrocław*, p. 87 (2007)

Assessment of Uterine Contractile Activity during a Pregnancy Based on a Nonlinear Analysis of the Uterine Electromyographic Signal

D. Radomski¹, A. Grzanka², S. Graczyk³, and A. Przelaskowski¹

¹ Division of Nuclear and Medical Electronics, Institute of Radioelectronics Warsaw University of Technology, Nowowiejska 15/19. 00-665 Warsaw, Poland

D.Radomski@ire.pw.edu.pl

² Institute of Electronic Systems Warsaw University of Technology, Nowowiejska 15/19. 00-665 Warsaw, Poland

Atekg@ise.pw.edu.pl

³ Department of Mother and Child Health, Poznan University of Medical Science, Polna 33, 60-535 Poznań, Poland

md-sg@neostrada.pl

Summary. Monitoring of a pregnancy course is one of socially important application of biomedical engineering in clinical medicine. In this paper we evaluated a possibility of a nonlinear analysis of an electrohystegraphical signal for assessment of an uterine contractile activity during a pregnancy. This analysis was performed based on a sample entropy statistic. The obtained initial results confirmed that this method could provide clinical useful information for an obstetrical care.

1 Introduction

Since many years information technologies have been supporting clinical medicine. Most of these applications are referred to cardiology, anesthesiology, neurology and radiology. On the contrary, reproductive medicine obeyed gynecology obstetrics and sexology is still mostly poor aided by information technology. This situation arises from two factors. Firstly, biological knowledge about physiology of human reproduction remains incomplete and hard to quantitatively modeling. Secondly, our culture gives impression that human reproduction process is highly mysterious and should evoke shyness. Thus, the serious challenge for the application of computer science to reproductive medicine is an observability problem, i.e. the problem how to measure biological features that determinate a state of a female or male reproductive system. In particular, high complexity of a female reproductive system makes that it can be unobservable globally.

The above limitations steer bioengineering researches to monitoring of a pregnancy course, particularly a last pregnancy trimester. This monitoring obeys observation of fetal well-being and observation of an uterine activity. It is performed using clinical methods such as palpable examinations of mother's abdomen

(e.g. Leopold's manipulation) and an assessment of a cervical opening process. The USG methods and biochemical tests are also performed. The primary drawback of these methods is impossibility of continuous monitoring of pregnancy and labor progress. Thus, they are complemented by a cardiotocography which enables to measure fetal heart rate variability and mechanical activity of an uterus.

Numerous results of epidemiological studies indicate that prolonged or preterm labors are the main risk factors of newborn neurological disorders [1]. Unfortunately, none of the above mentioned methods are effective clinically because they do not allow to predict beginning of a labor. Moreover, measuring of a mechanical activity of an uterus by one tocodynamic probe is biased by spatial-averaging of a measured signal. Additionally, this method provides only information on the frequency of the uterine contractions. It disables to estimate the force or efficiency of these contractions.

Overcoming those disadvantages, alternative methods are proposed. Among them the most promising is measuring of an electrical activity of an uterine muscle layer, called electrohysterography (EHG) [2]. The idea of this measuring is analogical to an electrocardiography. However, because of an electrophysiology of an uterus is still poorly known there is no elaborated standard method for analysis of EHG signals.

On the base of mathematical models describing a dynamic of uterine myometrium contractions and clinical observations some authors suggest that approaching labor contractions are evoked by a synchronization of electrical activities of myometrium fibers [3]. This hypothesis leads to appearance of higher regularity in EHG signals which may be identified by nonlinear signal analysis.

Although there are known many indexes which characterized nonlinear signal (e.g. fractal dimension, Lyapunov exponent etc), they are not sufficiently sensitive to signal regularity. Only the indexes constructed based on signal entropy have this required feature [4]. Graczyk *et al.* show effectiveness of an approximate entropy computed for EHG signal in assessment of a prelabor state of the uterus [5]. However, it was proven that the estimator of approximate entropy is biased and highly sensitive to number of signal samples [6]. Therefore, we study application of the unbiased entropy estimator called *sample entropy* for evaluation of the physiological state of the uterus.

2 Biological Basis for Pregnant Uterine Activity

Changes in uterine activity during a pregnancy is very complex process obeyed the interaction between an uterus and its cervix. This process is controlled by some hormonal, mechanical and electrical factors. The phenomenological model of this process is shown in Fig. 1 [3]. Uterine contractility is direct consequence of the electrical activity in the myometrial cells. An unpregnant uterus shows spontaneous electrical activities in the muscle from the uterus which are composed-intermittent bursts of spike action potentials. Uterine volume as well as ovarian hormones contributes to the change in action potential shape through their effect

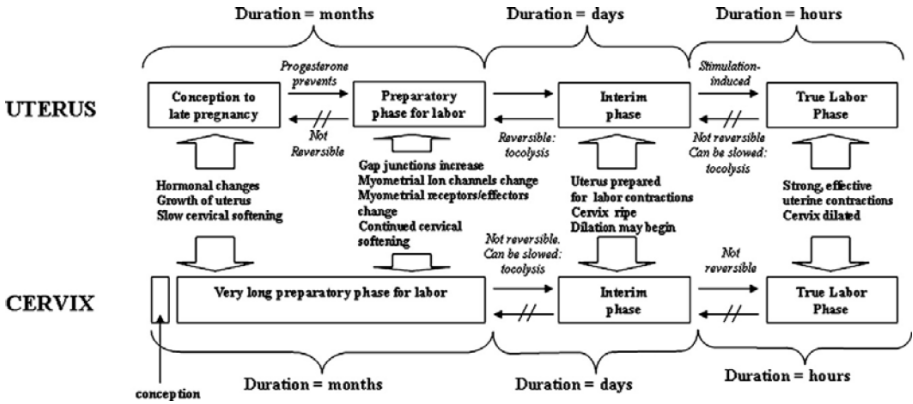


Fig. 1. A model of changes in an uterine activity during a pregnancy adopted from [3]

on resting membrane potentials. Usually multiple, higher-frequency, coordinated spikes are needed for forceful and maintained contractions [3]. The action potentials in uterine smooth-muscle result from voltage- and time-dependent changes in membrane ionic permeabilities. Ca^{2+} ions and Na^{+} ions contribute in generation of these potentials. The frequency, amplitude, and duration of contractions are determined mainly by the frequency of occurrence of the uterine electrical bursts, the total number of cells that are simultaneously active during the bursts, and the duration of the uterine electrical bursts, respectively. Each burst stops before the uterus has completely relaxed [3]. The potential wave is propagated in the uterus by so called gap-junction mechanism, i.e. myometrium cells are grouped and joined by channels with low electrical resistances. The electrical potentials propagated via myometrium cells can be observed in EHG signals as a spike complex. Generation of cellular potentials is well modeled by nonlinear dynamic model presented by Bursztyn *et al.* [7]. Thus, it can be considered that EHG signal is a time series generated by a nonlinear dynamical system.

3 Measurements and a Nonlinear Signal Analysis

3.1 Measurement of EHG

The analyzed signals were measured by the measuring system realized by ITAM Zabrze [8]. This system contains three channels: two for active Ag/AgCl electrodes and one channel TOCO probe. During a monitoring session, the electrodes were attached to the skin in the vertical median axis of the abdomen as it was shown in Fig. 2. The first EHG signals were measured differentially by the electrodes placed in the projection of body of the uterus and of the uterine cervix. The distance between the electrodes constituting the differential channels was set at 5 cm. The frequency range of EHG signal was limited to 5 Hz by low-pass filtration. The sampling frequency was 10 Hz.

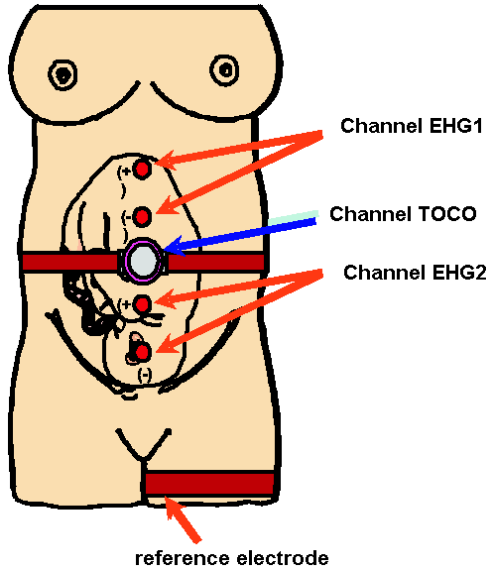


Fig. 2. The placement of the EHG electrodes and TOCO probe on female body

In the Fig. 3 is presented an example of the EHG signal registered by the EHG1 electrode and TOCO probe during a labor while Fig. 4 presents the EHG signal registered during a puerperium period.

3.2 Estimation of the Sample Entropy Statistic

Identification of EHG regularity was performed based on the sample entropy statistic. Let EHG signal will be represented by a time series denoting as $\{x(n)\}$. The sample entropy statistic was estimated in the following manner. Let's create m vectors contained consecutive values of x_i , commencing at the i -th sample, i.e.

$\mathbf{X}_m(i) = [x(i) \ x(i+1) \ \dots \ x(i+m-1)]$ for $1 \leq i \leq N - m + 1$. By $d[\mathbf{X}_m(i), \mathbf{X}_m(j)]$ is denoted the distance between two vectors $\mathbf{X}_m(i), \mathbf{X}_m(j)$ which is defined as:

$$d[\mathbf{X}_m(i), \mathbf{X}_m(j)] = \max_k |x(i+k) - x(j+k)| \tag{1}$$

The distance measure was used for counting of the number of the similar elements of the vectors $\mathbf{X}_m(i)$ and $\mathbf{X}_m(j)$. Let $\mathcal{J}_m = \{j : d[\mathbf{X}_m(i), \mathbf{X}_m(j)] \leq r\}$ will be the set of such indexes which for this distance is not greater than r . For a given \mathbf{X}_m and for $1 \leq i \neq j \leq N - m$ we define the coefficient $B_i^m = \frac{\text{card}\{\mathcal{J}_m\}}{N-m-1}$. Then, one can compute the number of the similar vector elements averaging over i , i.e. $B^m = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m$. It expresses the probability that two sequences coincide for m points. Analogically, such probability is computed for the vector \mathbf{X}_{m+1} . The estimator of the sample entropy is given by:

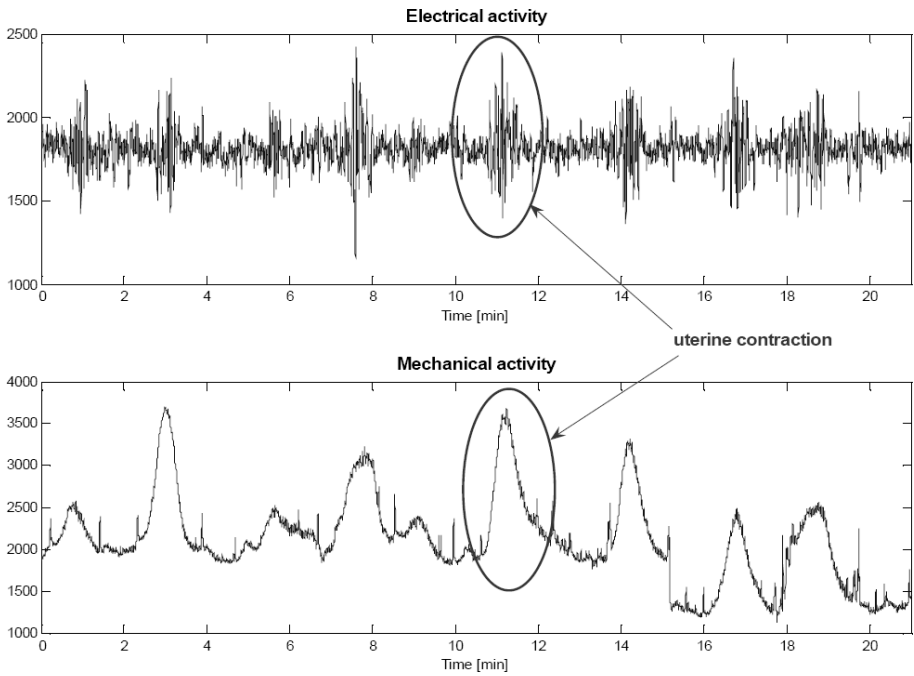


Fig. 3. An example of EHG and TOCO signal registered during a labor

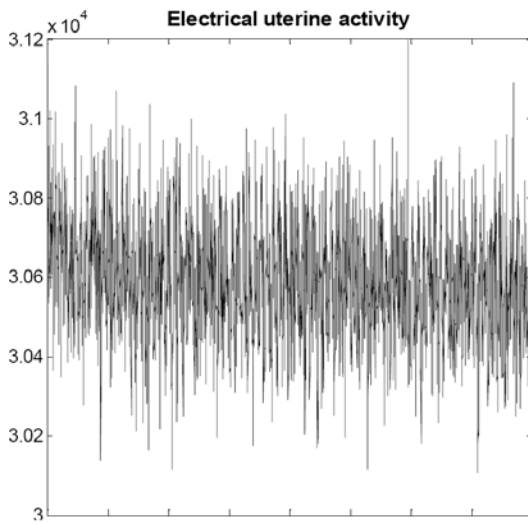


Fig. 4. An example of EHG signal registered during a puerperium period

$$\text{SamEn}(m, r) = -\ln \frac{B^{m+1}}{B^m} \quad (2)$$

This estimator depends on two parameters m, r . Setting their values properly is still problematic. In this study we assumed according to literature $m = 2, r = 0.2\sigma$, where σ was a standard deviation of a given EHG signal.

4 Results

The evaluation of EHG sample entropy for discrimination uterine activities was performed on the base of the signals registered from 66 pregnant women while 26 of them had registered EHG during term labors, 20 during preterm labors and 6 during puerperium periods. The mean values of the sample entropy estimators were computed. One-way ANOVA was used to evaluate differences between groups in these mean values. The results were shown in Fig. 5. All differences were statistically significant based on the LSD test.

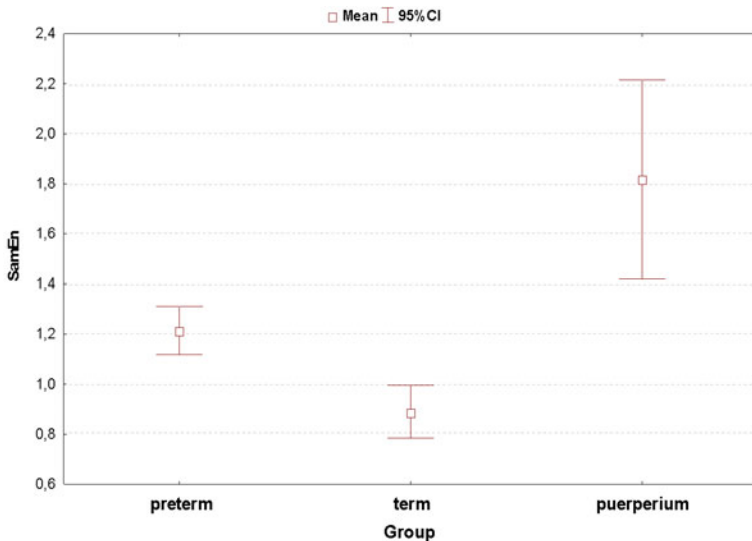


Fig. 5. The mean and 95% CI of the estimated values for EHG SamEn

5 Conclusion

High complexity of a physiological system controlling human partitions makes difficult to build phenomenological model of the electrical or mechanical activities of an uterus. However, there is a clinical need to monitoring them ensuring newborn health. The performed initial study shows that nonlinear analysis of EHG signals based on the sample entropy statistic could differentiate dynamic

states of an uterus. The sample entropy has several advantages over used the approximate entropy by other authors. This estimator reduces approximate entropy bias avoiding counting of self-matched elements. It leads to more consistent estimators [10] and decreases computing time consuming. This feature enables also a recursive form of SamEn estimators which could be used for 24 hours monitoring of an uterus activity [10]. Moreover, it is independent on time series length. However, impact of m, r parameters on discrimination power of the uterine states requires further investigations.

References

1. Goldenberg, R.L., Cliver, S.P., Bronstein, J., Cutter, G.R., Andrews, W.W., Mennemeyer, S.T.: Bed rest in pregnancy. *Obstet Gynecol* 84, 131–136 (1994)
2. Manner, L.W., MacKay, L.B., Saude, G.B., Garfield, R.K.: Characterization of abdominally acquired uterine electrical signals in humans, using a non-linear analytic method. *Med. Biol. Eng. Comput.* 44, 117–123 (2006)
3. Garfield, R.K., Manner, L.W.: Physiology and electrical activity of uterine contractions. *Semin in Cell Developmental Biol.* 18, 289–295 (2007)
4. Kanz, H., Schreiber, T.: *Nonlinear time series analysis*. Wiley, Chichester (2006)
5. Graczyk, S., Jezewski, J., Horoba, K., Wróbel, J.: Analysis of abdominal electrical activity of uterus - approximate entropy approach. In: *Proc 22nd Annual EMBS International Conference, Chicago* (2000)
6. Ferrario, M., Signorini, M.G., Magenes, G., Cerutti, S.: Comparison of entropy-based regularity estimators: application to the fetal heart rate signal for the identification of fetal distress. *IEEE Trans. Biomed. Eng.* 53, 119–125 (2006)
7. Bursztyn, L., Eytan, O., Jaffa, A.J., Elad, D.: Mathematical model of excitation-contraction in a uterine smooth muscle cell. *Am. J. Physiol. Cell Physiol.* 292(5), C1816–C1829 (2007)
8. Graczyk, S., Jezewski, J., Wrobel, J., Gacek, A.: Abdominal electrohysterogram data acquisition problems and their source of origin. In: *Proc. First Regional Conf. IEEE*, pp. PS13–PS14 (1995)
9. Aboy, M., Cuesta-Frau, D., Austin, D., Mico-Tormos, P.: Characterization of sample entropy in the context of biomedical signal analysis. In: *Proc. EMBS An Int. Conf.*, pp. 5942–5945 (2007)
10. Sugisaki, K., Ohmori, H.: Online estimation of complexity using variable forgetting factor. In: *Proc. SICE Annual Conference*, pp. 1–6 (2007)

Use of Computer System for Cell Hybridisation in Biotechnology and Medicine

Andrzej Dyszkiewicz^{1,2}, Paweł Połeć^{1,2}, Jakub Zajdel^{1,2},
Damian Chachulski^{1,2}, and Bartłomiej Pawlus¹

¹ Laboratory of Biotechnology Cieszyn ul.Goździków 2

² Computer Science Department, University of Silesia, 41 - 200 Sosnowiec
ul.Będzińska 36

Summary. An outline of issues relating to the contemporary application of monoclonal antibodies and techniques for obtaining hybrids has been presented in the study. The results of research concerning modifications of Zimmerman's methods have been presented. They prove a higher efficacy and selectivity of the solution proposed. An electroporation system with changed geometry of electrodes and current parameters has been presented. The procedure is controlled entirely by a microprocessor from the stage of technological parameters' control in the incubators of initial cells to the creation of a determined mixture of cells in a mixer, then division into portions, pumping the mixture into a hybridisation chamber and, following hybridisation, passing the mixture into separate sections of the incubator. The hybridisation system was modified. A transparent water coat was constructed and connected to a thermostat, on which a transparent hybridisation chamber was installed. Lighting from underneath and gap lighting of the chamber enable continuous observation of suspended cells by means of a microscope lens which is connected by a picture channel to a computer. The software analyses the picture in terms of hybrid selection. The marked cells are planimetrically analysed during the programmed duration. When the morphometric criteria are met, the suspended cells are pumped over to separate sections of the incubator, where selective breeding of hybrids is carried out. The selection of hybrids takes place in electroosmosis gradient under morphometric control of cells in microcapillary systems.

1 Introduction

One of the most significant problems connected with modern laboratory diagnostics and systems for selective control of medical therapy are monoclonal antibodies. They constitute a practical key to immunological and chemical peculiarity and present unique possibilities to find shield cells or even single chemical individuals selectively in organisms. The single and simplest antibody is tetramer, which is composed of two polypeptide heavy chains and two light chains connected by means of sulfhydryl bridges. The structure consists of an invariable area characteristic of a given immunoglobulin class and a variable area which consists of the ends of four chains, and their three-dimensional conformation and amino acid sequence almost fully determines their complete molecular peculiarity. The configuration of functional groups and unbalanced chemical bonds, as well as molecule vector decomposition of magnetic and electric fields, all in

the form of a three-dimensional mosaic are so unique that they are capable of determining the peculiarity of the molecule bond with a compatible chemical structure. At the same time, differentiation of chemical compounds takes place, not only on the basis of their molecular formulae, but also due to their three-dimensional conformation, including stereoisometry [20, 21]. Owing to a specific molecule identifier, diagnostics can link it with substances for radioactive, spin, fluorescent or classic dye marking. There also exist interesting therapeutic possibilities for this solution based on interconnection with cytostatic, antibiotic, or any drug, the performance of which should be very selective as its too high concentration in blood circulation can cause specified side effects in other organs [1, 2, 4, 5, 7, 8, 9, 10, 11, 14, 15, 16, 17, 18, 19]. Oncology seems to be a particularly fitting area where it can be applied.

- Precursor substances for monoclonal antibodies were plasmocytes – specialised cells of the immunological system, whose natural function was taking a selected antigen standard, e.g. from macrophages and production of antibodies compatible to its antigen determinants.
 - The next step was drawing attention to the occurrence of pathological plasmocytes showing an uncontrolled tendency to produce antibodies (plasmacytoma) in some types of growing diseases.
 - The basic biological structure capable of producing monoclonal antibodies are heterogen cell hybrids (atypical cells), which are created outside the myotic and meiotic cycle as a result of two initial cells merging, often of wholly different types. Because of numerous protection mechanisms such phenomena occur very rarely and spontaneously in natural conditions. It was observed that an inductive factor for such a change to occur may be viruses, multihydroxide alcohol, ionising radiation or electromagnetic fields. The above information was practically used in the design of systems for controlled hybridisation which employed:
 - Controlled infection of a determined solution of selective cell cultures for example with Sendai viruses
 - Exposure of the mixture of selective cell cultures to detergents which diminish surface tension of the cell's membrane /e.g. multihydroxide alcohol/
 - Direct exposure of the solution of selective cell cultures to electromagnetic fields and electric discharges /Zimmerman's method/ [12, 13, 23, 25]
- Nowadays, there exist many commercial laboratory systems (based on the Zimmerman notion) which enable conducting controlled hybridisation. Unfortunately, a common issue is their relatively low efficacy in obtaining living and dividable hybrids in comparison with the amount of cells which are put in for hybridisation and hybrids impaired and unable to divide any further.

The Aims of the Study

- computerised parameterisation of the hybridisation process in the extent of parameterisation of cell culture and dosage.

- controlling the proportion of the suspended matter loaded in which hybridisation is being carried out
- controlling and selection of hybridisation parameters
- continuous monitoring of hybrid forming dynamics

2 Characteristics of the System

The laboratory part of the system consists of an inspection chamber, situated under the microscope lens, from which, by means of a picture channel composed of a "Hiton" camera and "Fly video", converter card images of preparations of hybridised cells are inserted into the author's morphometric software design.

The hybridisation chamber is different from traditional solutions because of the geometric deformation of their electrodes[13]. The change in design leads to a greater predictability and efficacy of the hybridisation process because of specific guidance of the electric force field's line while cell cords are being formed and while needle potential takes place [23]. The hybridisation chamber is connected to a mixer chamber and into the latter determined amounts of cells are inserted from any number of thermostats by means of programmable peristaltic pumps. On filling the hybridisation chamber with a suspended matter of cells in specific proportions and on conducting electroporation controlled by the author's operating software, the cells are observed and judged by the author's morphometric software, which monitors the dynamics of their growth. If the sample obtained meets the morphometric criteria, the content of the chamber is pumped into a suitable incubator which enables further in vitro culturation.

Control of the Hybridisation Process

For control of the laboratory position, a 89C52 processor was used together with a multiplexor and drivers for specific functions of the system. Tool functions have been coded in a programmable memory of 32kB, there is an additional

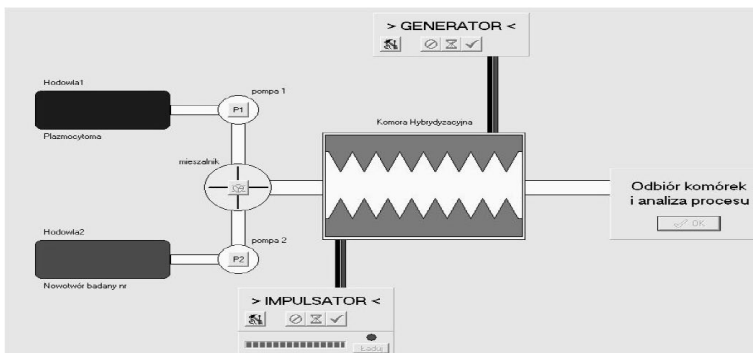


Fig. 1. Block diagram of the laboratory software for cell hybridisation

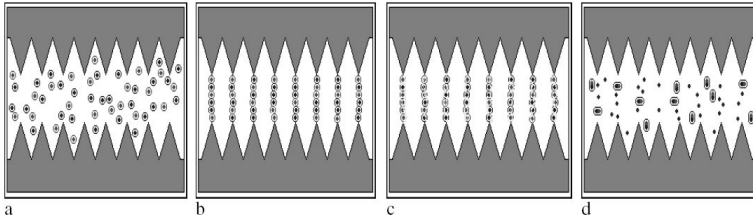


Fig. 2. Hybridisation phases: (a) introduction of a mixture containing a determined number of cells from the mixer; (b) actuation of the positioning field leads to the creation of interwoven chains of A and B cells; (c) generation of an electric impulse piercing the membranes of neighbouring cells; (d) pumping the contents of the chamber into a selective culture medium

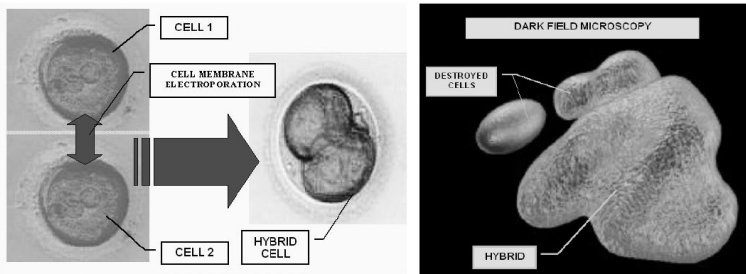


Fig. 3. Pictures and modeling of hybrid (fot. LaBio 2007)

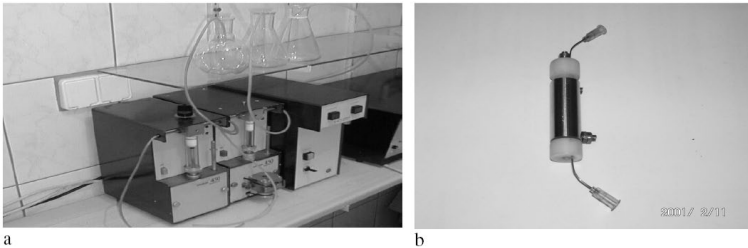


Fig. 4. Laboratory equipment (a) steering pumps; (b) hybridisation chamber

RS232 link for bi-directional HYBRID communication with a PC. The control of specific modules has been presented in detail in fig. 4.

3 Control Software

The software managing the system consists of two parts:

- electroporation module, written in "Delphi 2.0" which enables loading of voltage, intensity, frequency and the time of the current's passage in the

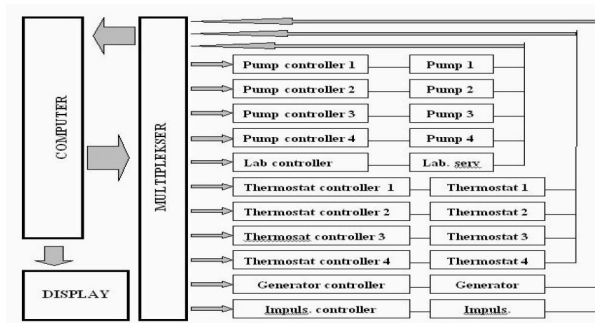


Fig. 5. Block diagram of the controller for parameters of the hybridisation process by means of a modified method of electrofocusing

hybridisation chamber, parameters steering the thermostat functions in a given culture, pumps proportioning the flow of suspended matters of cells and a cell mixer. The program co-operates with an experimental chamber in order to create optimal hybridisation parameters for a given hybrid type, and then, on the basis of the parameters obtained, it can control the parameters of the technological process of antibody production using the large hybridisation flow chamber.

- morphometric module, written in "C++" which enables loading microscope images from the hybridisation chamber (fig. 6) continuously or optionally. The program enables marking the chosen cell structures by use of a cursor, planimetric measuring of the cells marked at specific time intervals and generation of a diagram presenting changes in the cell area as a time function. The program enables also the creation of cell standards, which qualifies the hybrid obtained for further culturation.

Image Analysis Procedure

The registered image is loaded into the computer and then undergoes filtration by means of a median filter, which compared to a convolutional filter does not cause

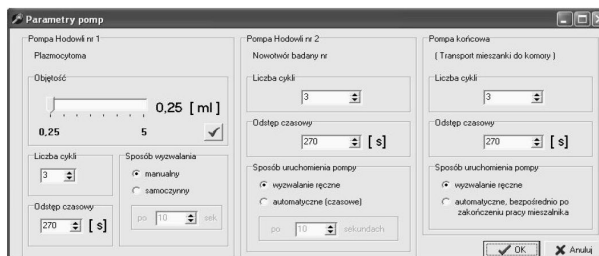


Fig. 6. Operation window for control of electrofusion parameters

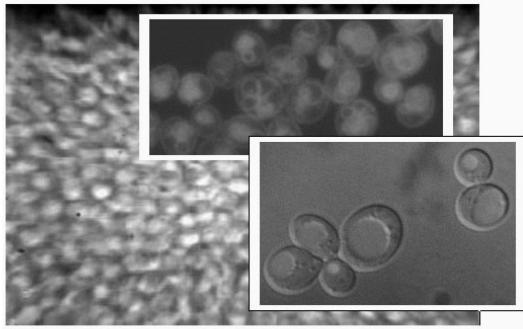


Fig. 7. Microscopic images of a cell suspension after leaving the hybridisation chamber at various magnifications

any noise drifting on a larger planar surface. The next stage involves histogram compensation i.e. converting the brightness of particular image points so that the number of points whose brightness lies within equal range in the histogram is the same.

As can be seen in the picture (fig. 7), the projection of many bodies onto the surface leads to the creation of a preconjugated system where it is difficult to identify the base components. Assuming a relatively large loss of information, the components may be separated by controlling the binarisation threshold (fig. 8). Thus for binarisation with a low threshold, the subsequent opening and closing operations involve elementary morphological conversions such as are erosion and dilatation or dilatation and erosion in the order provided.

Erosion involves removing all the points of an image with a logical value of 1 which are neighboured by at least one value of 0. In the dilatation process the value of the central point is 1 on the condition that not all the values of the neighbouring pixels are equal to 0. Both erosion and dilatation generalize the image by removing small concavities - dilatation, or small, isolated areas

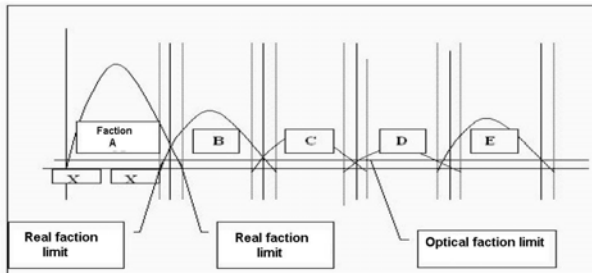


Fig. 8. Presentation of the conjugation of overlapping projections of cell bodies onto a plane with clearly visible conjugation areas A/B; B/C; C/D; D/E

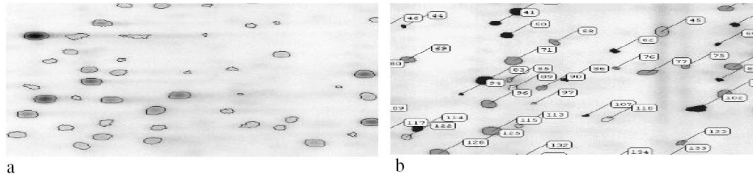


Fig. 9. Following binarisation which guarantees the separation of the cell structures' image, the following procedures can be carried out: (a) outlining; (b) indexation

- erosion. The described procedures, if used separately, lead to a change in the surface area of the significant diagnostic areas.

The obtained image, void of minute interference, is closed by an edge algorithm and then is subject to indexation i.e. assigning determining identifiers to all the pixels belonging to the structures, indicating which structure a given pixel belongs to. This operation allows for the exposition of all the fragments of an image which are of interest, as well as for removing interference by introducing a minimum threshold for the size of the structure.

Observation of a Structure

Planimetric assessment of the dimensions of static structures is not much of an issue, especially when having clear edge parameters in the matrix which have been compared with standard metric units. Assigning planar parameters to pixel sets enclosed in edge contours does not cause much difficulty either (fig. 10a). The situation becomes much more complex when observing live structures. The overlapping of metabolic micromotions, chemotaxis and excretory functions leads to the observation of edge contour drift (fig. 10b) during continuous recording of images in a gate time approximate to the micromotion function duration.

Moreover, a cell in a water solution has limitless freedom of movement hence its 2-dimensional projection enters into unpredictable conjunctions with other structures in subsequent images. In such conditions, the specificity of the algorithm for tracking the indexed cells based on their edge contour falls dramatically.

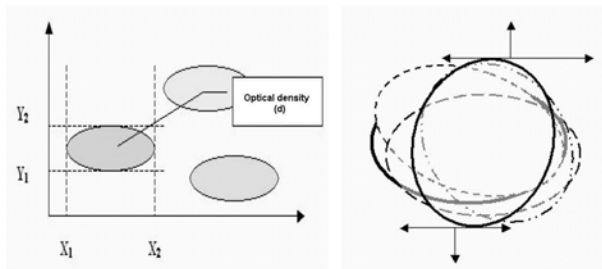


Fig. 10. Planimetric positioning of index numbers based on the edge contour or the structure image's center of gravity

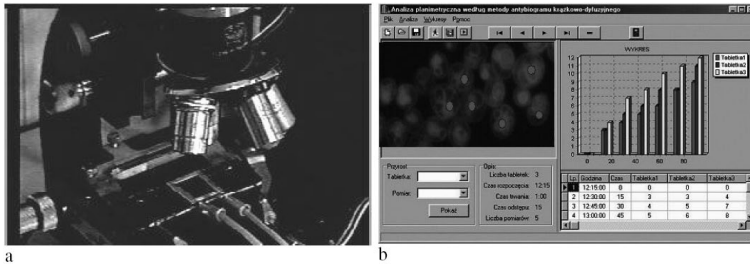


Fig. 11. Observation of the growth of the indexed structures' edge contours (growth of cell hybrids): (a) Microscope chamber for cell control, (b) software window for measurements (fot. LaBio 2007)

This algorithm, however, has a very valuable feature which allows for a comparison to be made of the cell's dimensions at regular intervals which provides the possibility to assess its growth dynamic.

While considering the issues which have been presented above, the measurement process of the growth of marked cells can be seen as a compromise between the capabilities of laboratory preparation and information technology methods. The preparation process is aimed at achieving maximum immobilisation of hybrids on semi-fluid or solid surfaces by operating with surface tension parameters, oxidation-reduction potential and by selecting their optimum chemical composition.

Conclusions

The laboratory position presented is a modern system, which combines the practical capabilities behind many interesting biotechnological solutions. The system is a quick compromise between demand, dictated by the market, and production potential. It enables an easy specification of biotechnological parameters in the production process to be made of any type of hybrid. This is possible thanks to the parameter transformation of an experimental chamber into a technological chamber. By means of computer controlled hybridisation, steering systems and storing cell images, it is possible to create interactive and quickly accessible database systems. The system may be very useful in long term production programs. However, it is cardiology [3, 6], endocrinology [4, 21] and oncology which seem to be a particularly interesting area for its implementation.

References

1. Badkar, A., Betageri, G., Hofmann, G.: Enhancement on transdermal iontophoretic delivery of a liposomal formulation of colchicine by electroporation. *Drug. Deliv.* 6, 111 (1999)
2. Banga, A., Prausnitz, M.: Assessing the potential of skin electroporation for the delivery of protein and gene based drugs. *Trends in Biotech.* 16, 408 (1998)

3. Banić, A., Brewer, L., Wendt, M.: Local delivery of platelets with encapsulated iloprost to balloon injured pig carotid arteries effect of platelet deposition and neointima formation. *Thromb. Haemost.* 77(1), 190 (1997)
4. Chang, S., Hofmann, G., Zhang, L.: Transdermal ionophoretic delivery on salmon calcitonin. *Int. J. Pharmac.* 200, 107 (2000)
5. Chazal, M., Benchimol, D., Bernard, J.: Treatment of liver metastases from colorectal cancer by electrochemotherapy. In: *Annual Cancer Symposium Society of Surgical Oncology*, vol. 35, p. 18 (1998)
6. Cui, J., Robinson, K., Brown, J.: Local drug delivery to pig carotid arteries by direct vessel wall electroporation using a novel catheter. In: *Proceedings of the American College of Cardiology*, vol. 29(2), p. 201A, 749 (1997)
7. Dev, S., Giordano, F., Brown, D.: Can endoluminal gene delivery by pulsed electric field overcome the current deficiencies of gene therapy for cardiovascular diseases. *Proceed. Bioelectrochem. Bioenerg.* 113 (1996)
8. Dev, S., Hofmann, G.: Electrochemotherapy a novel method of cancer treatment. *Cancer. Treat. Rev.* 20, 105 (1994)
9. Dev, S., Nanda, G.: Electrochemotherapy- what does the future hold? Results from treatment of human pancreatic and non-small cell lung cancer xenografted onto nude mice. *Proceed. Bioelectrochem. Bioenerg.* 143 (1996)
10. Dev, S.: Electrochemotherapy for cancer. *Cancer Watch.* 5(2), 23 (1996)
11. Dev, S.: Killing cancer cells with a combination of pulsed electric fields and chemotherapeutic agents. *Cancer Watch.* 3, 12 (1994)
12. Dyszkiewicz, A.: Komora do elektrofonoforezy. *Urząd Patentowy RP*, 10/1998, W 108753
13. Dyszkiewicz, A.: Komora do hybrydyzacji komórek oraz diagnostyki płynów ustrojowych. *Urz Pat RP P* 328254
14. Dyszkiewicz, A., Gaździk, T., Barańska, T.: Drug penetration into muscle tissue after phonophoresis, ionophoresis and electrophonophoresis. *Acta Bioeng. Biomech.* 1 (suppl 1) 125 (1999)
15. Dyszkiewicz, A., Sapota, G., Wróbel, Z.: Wielofunkcyjne sterowanie procesem elektrofonoforezy w terapii zespołów bólowych kręgosłupa. In: *II Sympozjum MPM Krynica Górská 8-12.05.* p. 1311(2000)
16. Dyszkiewicz, A., Imielski, K.: Kliniczne a laboratoryjne hodnoceni penetrance leku w procesie elektrofonoforezy. *Rehabil Fizik Lekar*, 4 (2000)
17. Dyszkiewicz, A., Gaździk, T.: Ocena penetracji leków w tkance mięśniowej po zastosowaniu elektrofonoforezy. *Post Rehab*, 4 (1999)
18. Dyszkiewicz, A., Wróbel, Z.: Elektromagnetoforeza presyjno-rotacyjna. *Urz Pat RP*, 01/ 2001
19. Dyszkiewicz, A., Wróbel, Z.: Reversible remote modification of the skin penetration for ions and particles based on the pressure-rotation electromagnetophoresis. *Int Symp Biophys Med, Cieszyn* (2001)
20. Eisenbarth, G.: Application of Monoclonal antibody techniques to biochemical research. *Anal. Biochem.* 3, 1-16 (1981)
21. Eisenbarth, G., Jackson, R.: Application of monoclonal antibody techniques to endocrinology. *Endoc. Rev.* 1, 26 (1982)
22. Glass, L., Pepine, M., Fenske, N.: Bleomycin mediated electrochemotherapy of metastatic melanoma. *Arch. Dermatol.* 132, 1353 (1996)
23. Heller, R., Jaroszeski, M., Atkin, A.: Electrically enhanced delivery of molecules to cells. In: *Proceedings of the Congress on in Vitro Biology* (1997)

24. Heller, R., Jaroszeski, M., Perrott, R.: Effective treatment of B16 melanoma by direct delivery of bleomycin using electrochemotherapy. *Melanoma Research* 7, 10 (1997)
25. Hofmann, G., Dev, S., Nanda, G.: Electrochemotherapy- transition from laboratory to the clinic. *IEEE Eng. Med. Biol.* 15(6), 124 (1996)
26. Hofmann, G., Rustrum, W., Suder, K.: Electro-incorporation of microcarriers as a method for the transdermal delivery of large molecules. *Bioelectrochem. Bioenerg.* 38, 209 (1995)
27. Nishi, T., Dev, S., Yoshizato, K.: Treatment of cancer using pulsed electric field in combination with chemotherapeutic agents or genes. *Human cell* 10(1), 81 (1997)
28. Sersa, G., Kranjc, S., Cemazar, M.: Improvement of combined modality therapy with cisplatin and radiation using electroporation of tumors. *Int. J. Radiat. Oncol., Biol., Phys.* 46(4), 1037 (2000)
29. Widera, G., Austin, M., Rabussay, D., Increased, D.N.A.: vaccine delivery and immunogenicity by electroporation in vivo. *J. Immunol.* 164, 4635 (2000)
30. Widera, G., Dev, N., Nolan, E.: Immune responses after DNA vaccination augmented by in vivo electroporation. In: *Keystone Symposium, DNA Vaccines; Immune Responses; Mechanisms and Manipulating Antigen Processing*; 12-17.03, Snowbird, UT0 (1999)
31. Zhang, L., Hofman, G., Li, L.: Transderma delivery of genes by pulsed electric fields. *Proceedings Bioelectrochem. Bioenerg.* 147 (1996)
32. Zhang, L., Li, L., An, Z.: In vivo transdermal delivery of large molecules by pressure-mediated electroincorporation and electroporation; A novel method for drug/gene delivery. *Bioelectrochem. Bioenerg.* 42(2), 283 (1997)

Clustering as a Method of Image Simplification

Anna Korzynska and Mateusz Zdunczuk

Laboratory of Microscopic Image Processing Information Systems, Department of Hybrid Biosystems Engineering, Polish Academy of Sciences Institute of Biocybernetics and Biomedical Engineering, 4 Trojdena Str., 02-109 Warsaw, Poland
akorzynska@ibib.waw.pl

Summary. The microscopic images of the cells are very difficult to analyze and to segment. The advanced method of segmentation such as region growing, watershed or snake requires the initialization information about the rough position of the cell body. It is proposed to localize cells in image using a threshold of simplified image. Clustering grey levels in image is proposed to simplify image. The k -means clustering method supported by weighting coefficients is chosen to collect all grey tones presented in the background into one cluster and other grey tones into few clusters in such a way that they cover a cell region in microscopic images. The weighting coefficients are used to influence (expand or contract) patterns in microscopic images of living cells. The method was evaluated on the basis of confocal and bright field microscopy images of cells in culture.

1 Introduction

The microscopic images of the cells are very difficult to analyze because of lack of precise and accurate methods of cells separation from the background. The segmentation of cell images are not easy due to the contrast quality, variation in cell shape, temporal changes in image contrast and focus, what is shown in Fig. 1.

The advanced method of segmentation such as region growing, watershed or snake requires the initialization information about the rough position of the cell body. The main idea of this research is to localize cells in image using one of the clustering methods. The k -means clustering method supported by weighting coefficients is proposed to reduce quantity of grey tones in image in such a way that the background variation is suppressed into one cluster and the other clusters cover cell area. This type of simplified image, after being thresholded, allows to localize the cell body fragments. Next, using mathematical morphology, binary operations, the cell fragments would be connected and holes in their area would be filled.

2 Related Research Review

Clustering has a rich history in pattern recognition [29], image processing [6, 10] and information retrieval [15, 16]. In this paper this methodology is employed in image processing as a low-level procedure that aims at simplifying an image.

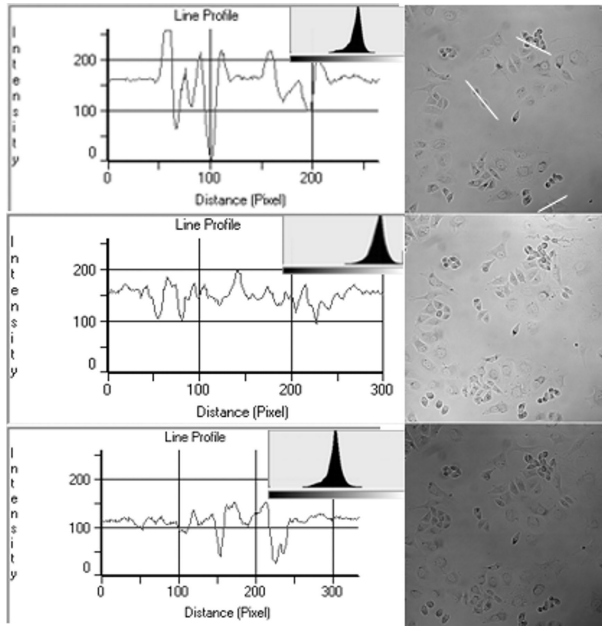


Fig. 1. Three images with their histograms present external source light variations and internal nonhomogeneity in the light distribution. There are three line profiles: the top line profile shows intensity function along the top white line crossing the relatively well contrasted and small cell cluster, the central one shows intensity function along the middle white line crossing the middle size cells, and the last one shows intensity function a long bottom white line crossing large, flattened and poorly contrasted cell.

Analogical, but not the same aims have been investigated by Du et al. [8] and by Duda and Hart [9]. Both groups of researchers were concentrating on partitioning an image into homogeneous regions in the sense of segmentation rather than object localization. Among methods of image segmentation, using clustering, the most interesting are: k -means [13, 25], isodata [3], and fuzzy c -means [28]. In this investigation the variation of k -means method was chosen as very easy to adjust to particular image by manipulation of weighting coefficients.

The k -means clustering methods were introduced in 1967 by J. MacQueen as an unsupervised classification technique. These methods were used to detect cell nuclei in digital image-based cytometry [20, 26], but only for fluorescent microscopic images which are easier to analyze rather than bright field or confocal microscopy images. Since bright field and confocal microscopic images segmentation using such methods as watershed [21, 17], region growing [18, 24], model-based [22] and agent based or hybrid method [2, 4], gives good accuracy and precision but all these methods need initial information of the cell position, so the k -means clustering is proposed to be pre-processing phase of these images segmentation methods.

3 Microscopic Image Characterization

The microscopic images of cells are very difficult to analyze. This is because of the image quality which depends on a type of microscope, a type of cells and a resolution of acquired image [19, 7].

The bright field microscopy and the scanning confocal microscopy produce greyscale images with poorly contracted cells in the image plane, see Fig. 1. Cells are transparent objects so they transmit light and they are visible as grey tones which are darker or brighter than grey background. Some part of the cell body is in the background grey level range. Some cells are partly or fully rounded by halo which appears as brightened background. This brightness is caused by light reflection on cell wall. There is no halo and contrast between a cell and the background in the parts where cell is very flattened. Furthermore the intensity of the background is not uniform across the image, due to the external and source light variation.

Three graphs of line profiles in Fig. 1 present significant intensity changes across the image plane (bottom-right and top-left image corners are darker than bottom-left and top-right) in both, in the background and within the cell. The variations of the intensity caused by noise is also observed in the background and in the cell area. The histogram of each microscopic cell image is unimodal and slightly skew. The presented histograms are located in various positions of grey scale according to mean lighting conditions, see Fig. 1.

There is a difference in detail visibility according to microscopic techniques. Neural stem cells sample observed in the red laser confocal scanning microscopy and the bright field microscopy image are presented in Fig. 2. The bright field microscopy builds images (see right part of Fig. 2) with relatively large deep of field in comparison to the laser confocal microscopy (see left part of Fig. 2).

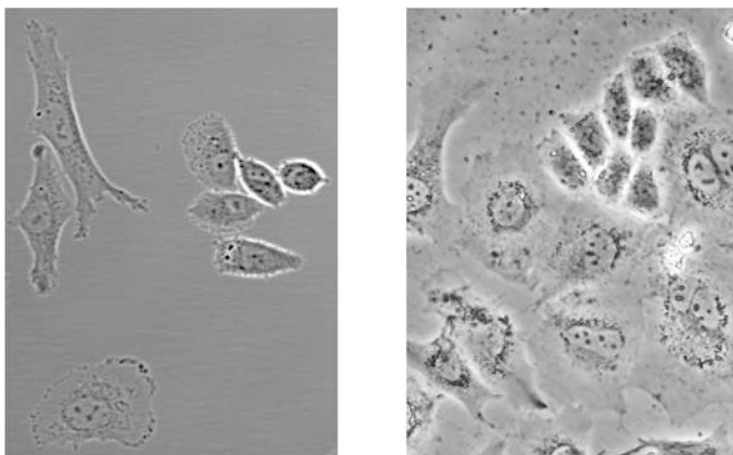


Fig. 2. Cells images in the red light laser scanning confocal microscopy (left) and the bright field microscopy (right)

The structures visible in confocal microscopy such as nucleus, the endoplasmic reticulum around nuclei only sometimes are observed in bright field images as blurred objects.

4 Methods

The goal of the study is to find the method of the microscopic image simplification which suppress the background variation into one grey tone cluster and cell regions in the other clusters. It is proposed to use k -means clustering method to do that.

4.1 Clustering Method

Clustering is a commonly used technique for features determination and extraction from large data sets and for the determination of similarities and dissimilarities between elements in the data sets.

In this paper the concept of image clustering is exploited. Image pixels are grouped in such a way that the information which image contains is emphasized or extracted. Generally, the criteria used to collect pixels in clusters are various, among them color, texture, connectivity, gradient and so on and they depend on image characteristics and objects in the image and on the goal of the image processing. Pixels collected in one cluster are presented by similar or the same color or grey level in simplified image what leads to image segmentation. This process reduces image noise and some image details. If these details carrying information which is redundant or not important according to the goal of the image processing, the image simplification does not damage image, but rather highlights its sense.

Mathematical Background

The clustering problem can be formally defined as follows [12, 27].

Given a data set $U = \{u_1, u_2, \dots, u_p, \dots, u_{N_p}\}$ where u_p is a pattern in the N_d -dimensional feature space, and N_p is the number of patterns in U , then the clustering of U is the partitioning of U into K clusters $\{V_1, V_2, \dots, V_K\}$ satisfying the following conditions:

- Each pattern should be assigned to a cluster, i.e.

$$\bigcup_{k=1}^K V_k = U$$
- Each cluster has at least one pattern assigned to it

$$V_k \neq \emptyset \text{ for } k = 1, \dots, K$$
- Each pattern is assigned to one and only one cluster

$$V_l \cap V_k = \emptyset \text{ for } l \neq k.$$

Clustering can be defined with reference to an image [8, 23], than given an image u , let $U = \{u(i, j)\}_{(i, j) \in D}$, where $D = \{(i, j) : i = 1, \dots, I, j = 1, \dots, J\}$

for positive integers I and J , (i, j) are integer pairs that range over the image domain, denote the set of grey tones values in the original image. Then, for any set of the replacement grey tones $W = \{w_k\}_{k=1}^K$, called generators, let

$$V_l = \{u(i, j) \in U : |u(i, j) - w_l| \leq |u(i, j) - w_k|, k = 1, \dots, K\} \quad (1)$$

$V_l, l = 1, \dots, K$ denotes the subset of those values of the grey tones that are closest to w_l (in the sense of the euclidian distance in 1-dimensional space of grey levels) than to any of the other w_k 's. The subset of grey tones V_l is called the cluster corresponding to w_l and the set of subsets $V = \{V_k\}_{k=1}^K$ is called a clustering of the set U of grey tones. For any non-overlapping covering of $U = \{V_k\}_{k=1}^K$ into K subsets, one can define the means or centroids of each subset V_k as the grey value $\bar{w}_k \in V_k$ that minimizes following expression called clustering energy

$$\sum_{k=1}^K \sum_{u(i,j) \in V_k} |u(i, j) - w_k|^2 \quad (2)$$

The grey values w_k that generate the clustering such that $w_k = \bar{w}_k$ for $k = 1, \dots, K$ are called the centroids of the associated clusters.

Several variants of the k -means algorithm have been reported in the literature [1]. Some of them attempt to select a good initial partition so that the algorithm is more likely to find the global minimum value. Another variation is to permit splitting and merging of the resulting clusters. Typically, a cluster is split when its variance is above a pre-specified threshold, and two clusters are merged when the distance between their centroids is below another pre-specified threshold. Using this variant, it is possible to obtain the optimal partition starting from any arbitrary initial partition, provided proper threshold values are specified. Another variation of the k -means algorithm involves selecting a different criterion function. The weighted k -means algorithm is a variation of the classic k -means algorithm [14, 29]. Weight coefficients, which provide weighted distortions between data and cluster centers, are incorporated into the algorithm to realize anticipated clustering. One can redefine the energy expression Eq. 2 so that the contributions from each of the clusters are weighted. This allows, for example, for a given grey tone to be included in a large cluster and opposite. Applied to a digital image, weighted clustering can let grey tone generators focus on selected details of the image and not be overwhelmed by other grey tones. Definition of the weighted energy expression is as follows

$$\sum_{k=1}^K \lambda_k \sum_{u(i,j) \in V_k} |u(i, j) - w_k|^2 \quad (3)$$

where λ_k are positive weighting factors. In general, λ_k is allowed to depend on factors such as the cardinality $|V_k|$ of the subset V_k , the within cluster variance, etc. [14]. In the resulting image original grey levels are replaced by centroids of the last iteration of the proposed algorithm.

Algorithm

The k -means method aims to minimize the sum of squared distances (in the sense of grey levels) between all points and the cluster center. Algorithm works recursively:

1. Initialization phase:

- a) Choose K initial cluster centers w_1, w_2, \dots, w_K .
- b) At the k -th iterative step, distribute the samples $u(i, j)$ among the K clusters using the relation,

$$u \in V_k \text{ if } |u(i, j) - w_l| \leq |u(i, j) - w_k| \text{ for } k = 1, \dots, K \quad (4)$$

where V_k denotes the set of samples whose cluster center is w_k and $l \neq k$.

2. Iterative phase

- a) Compute the new cluster centers $w_k(k+1)$, $k = 1, \dots, K$ such that the sum of the squared distances from all points in V_k to the new cluster center is minimized. The measure which minimizes this is the sample mean value of V_k . Therefore, the new cluster center is given by

$$w_k(k+1) = \frac{1}{N_k} \sum_{u \in V_k} u \text{ for } k = 1, \dots, K \quad (5)$$

where N_k is the number of samples in V_k .

- b) If $w_k(k+1) \cong w_k(k)$ for $k = 1, \dots, K$ (with respect to a chosen threshold value) then the algorithm has converged and the procedure is terminated. Otherwise go to step 2.

Different stopping criteria can be used in an iterative clustering algorithm:

- the change in centroid positions are smaller than a user-specified value,
- the quantization error is small enough,
- a maximum number of iterations has been exceeded.

In the proposed method the last stopping criterion is used and merging procedure supported by weighting coefficients are exploited. So the resulting cluster consists of clusters corresponding for grey levels of the background. The weighting coefficients are selected on the basis of the cluster size counted from the previous iteration.

4.2 Details of the Proposed Method

A major problem with k -means algorithm is its sensitivity to the selection of the initial generator number and their positions and its convergence to a local minimum of the criterion function if the initial partition is not properly chosen. In this investigation the number of the generators ranges from 5 to 25 and three various methods of choosing the initial generators were tested:

1. random choice,
2. homogenous choice,
3. arbitrary choice,

of the grey levels over grey scale. Resulting images for various values of K and for various methods of generators initialization are presented in Sect. 6.

5 Material

Evaluation of the proposed method was done using microscopic images of neural stem cell culture [5]. The images were acquired from digital cameras attached to two microscopes: bright field inverted microscopy (Olympus IX70 the right side Fig. 2) and scanning confocal microscope with the red color laser (Zeiss Fig. 1 and the left side Fig. 2). The observation plane on culture dishes cover $100 \times 100 \mu\text{m}$ space in the first case while $120 \times 120 \mu\text{m}$ in the second. Bright field microscopy images were acquired as a digital image of 1024×1024 pixels in 12 bits deep acquisition and next converted to 8-bit deep images by the linear resampling of the grey scale from minimum to maximum of grey levels. In the case of confocal microscopy images 8 bits deep acquisition of 2048×2048 pixels were done. Bicubic resampling was used to resample images to the size 1024×1024 .

6 Results

The proposed method results on microscopic images were analyzed and compared in both qualitative and quantitative manner.

6.1 How Initialization Influences the Results

Fig. 3 shows the results of the proposed method with increasing number of clusters and with various generators distribution over grey scale.

It can be observed that the number of clusters influences the resulting image in the detail level and in the smoothing of the background. The larger cluster number, the more grey values are bounded up with the background. Therefore smoothing of the background is desirable, the choice of 8 clusters seems to be the best to achieve the smoothing of the background. The dependence of the results on the way generators are chosen isn't unambiguous. Because tested strategies of generators choice do not lead up to fundamental differences in the final distribution of the centroids after many iterations and difference among results are not ambiguous and not significant, arbitrary choice of generators positions was used in further investigation.

6.2 Evaluation of Influence of Selected Weighting Coefficients

The procedure that involves weighting coefficients aims at such partitioning of the picture grey levels that some pixels which grey tones correspond to the

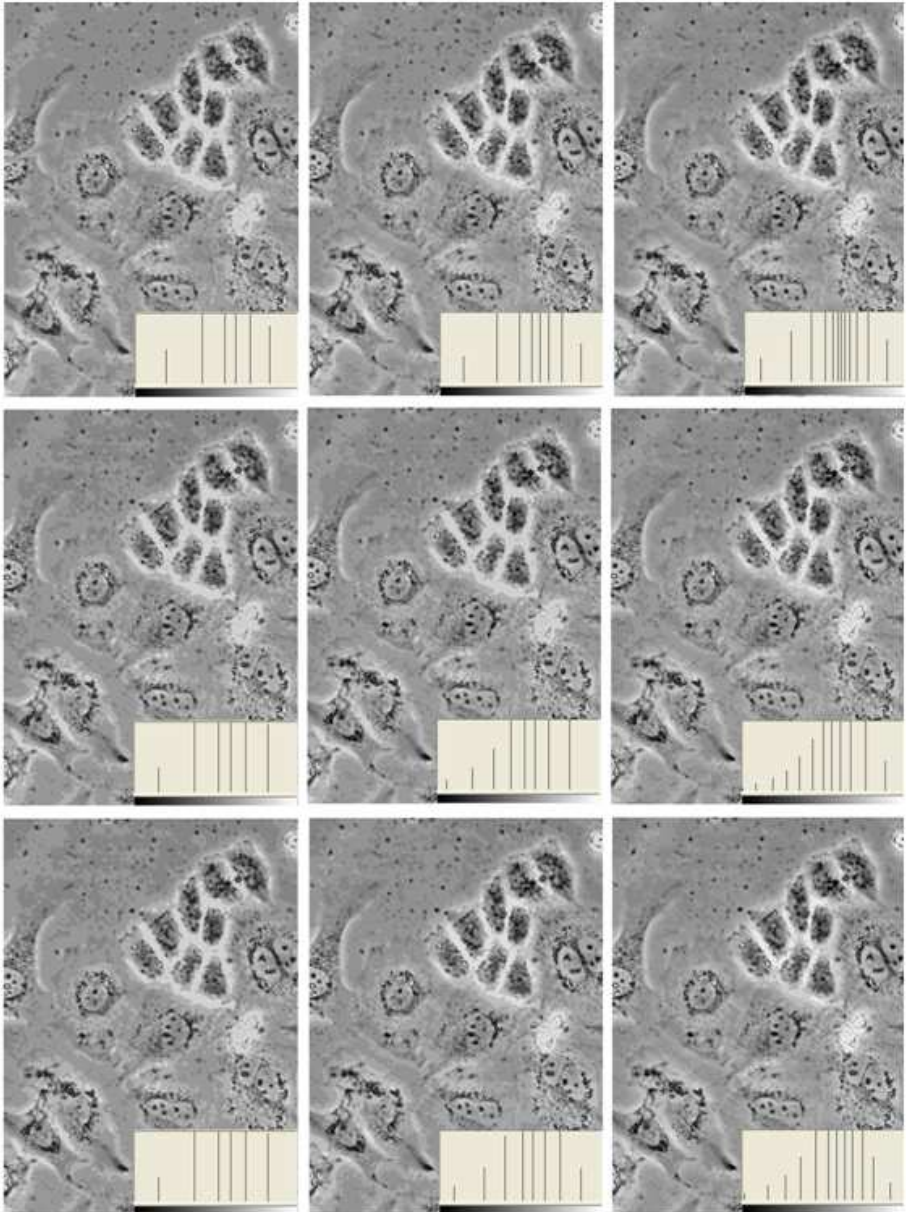


Fig. 3. Proposed clustering method results with corresponding image histograms for various initialization methods: with increasing number of clusters from $k = 6$ in the first column, by $k = 8$ in the second one to $k = 12$ for the last one and with starting centroids chosen as: randomly distributed across the grey scale in the first row, homogeneously distributed in the second one and arbitrary chosen by operator in the third one

Table 1. Matching matrix resulting from weighted k -means algorithm: distribution of classified pixels percentages compared with classification without weighting coefficients

Image	Weighting coefficients	Centroids of the last iteration; Resulting centroids	Fraction of pixels displaced
1-row,1-column	0 0 0 0 0 0 0	32 73 98 108 119 140 177 239	—
2-row,1-column	150 0.01 0.01 2048 10456 5586 0.01 45	43 74 85 113 141 185	81,2%
2-row,2-column	150 654 856 0.01 0.01 0.01 75 130	43 74 85 95 113 141 245	45,7%
2-row,3-column	150 654 856 2048 10456 5586 75 130	43 74 85 95 113 141 185	62,1%

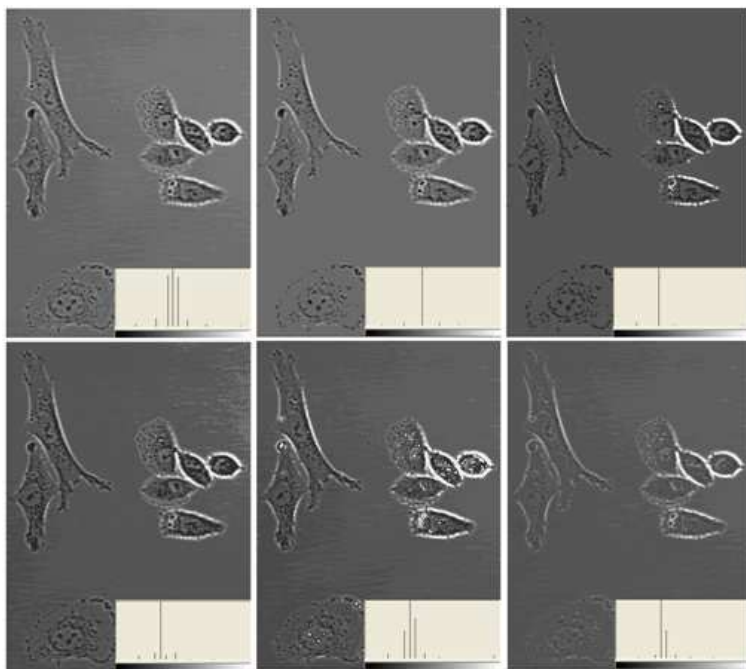


Fig. 4. The results of the proposed clustering method for arbitrary chosen generators position as described in the text with corresponding image histograms: - in the first row (from left to right): reference image with all weighting coefficients equally influences all clusters (equal 1); reference image with merged background grey levels (98 and 119 merged to 108); image with weighting coefficients which influence five first clusters [0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 1, 1]; in the second row images with coefficients calculated based on cardinality of the chosen clusters (from left to right): resulting in the homogeneity of the background, resulting in the expansion of the objects, resulting in both homogeneity of the objects and homogeneity and the expansion of the background

one class are placed in the other class. Generally, pixels are attracted by other classes if coefficients of these classes are small numbers. This mechanism is used

to include background pixels around cells structures to the cell classes to achieve a cohesive object region. Matching matrix, presented in Tab. 1, was calculated to evaluate fraction of pixels reclassified from one class to the other one according to a chosen set of coefficients. Classification without weighting coefficients, which is presented in both the first row in the Tab. 1 and in Fig. 4 the first row and the first column image, is used as a reference to calculate pixels displacements. The last three rows of Tab. 1 present the matching matrix results for images shown in the second row in Fig. 4.

6.3 How Cardinal Coefficients Influences the Results

Fig. 4 shows the results of the proposed method with 8 clusters and arbitrary chosen positions of initialization generators: [42 68 86 109 137 149 162 175] and with various values of weighting coefficients. In this test weighting coefficients different than equal to 1 appeared for various clusters to investigate influence of specific cluster expansion or contraction in the resulting image.

It was observed that in order to narrow the area of the determined cluster, the coefficients are selected to be equal to the cluster size. Otherwise, in case of the cluster area expansion, coefficients are selected as the inverse of the cluster size. The results show that choosing such coefficients that are leading to the background expansion (first image in the second row), gives the best effects while the grey tones present in cell area are manipulated the results images are worse (middle and last images in the second row). The best result was obtained by merging procedure (middle image in the first row).

7 Discussion and Conclusion

The proposed greyscale image simplification method described in this paper employs clustering method to redefine grey value of each pixel in microscopic images in such a way that the background pixels are collected into one cluster and the other clusters collect pixels of other grey tones. Resulting images, presented in the text, show that the used method is less dependent on the way the starting centroids are chosen rather than on the number of these centroids and that weighting coefficients based on the cluster cardinality allow manipulation of the cluster size. These studies conclude that the proposed method is promising enough to carry out the influence on the use of another types of coefficients. It seems that texture would be a good choice for a new coefficient definition because texture is the feature which discriminate area of the cell from the background area.

Acknowledgement. We are grateful for cell impart from line HUCB-NS to the experiments and for the support in experiments received from the NeuroRepair Department Laboratory, Polish Academy of Sciences Medical Research Center.

References

1. Anderberg, M.: *Cluster Analysis for Applications*. Academic Press, New York (1973)
2. Baujard, O., Garbay, C.: KISS: a multiagent segmentation system. *Optical Engineering* 32(6), 1235–1249 (1993)
3. Bezdek, J.C.: A Convergence Theorem for The Fuzzy ISODATA Clustering Algorithms. *IEEE Transaction On Pattern Analysis And Machine Intelligence* 2(1), 1–8 (1980)
4. Boucher, A., Doisy, A., Ronot, X., Garbay, C.: Cell Migration Analysis Afte. *Vitro Wounding Injury with a Multi Agent Approach*. *Artificial Intelligence Review* 12, 137–162 (1998)
5. Buzanska, L., Jurga, M., Stachowiak, E.K., Stachowiak, M.K., Domanska-Janik, K.: Focus on Neural Stem Cells. Neural Stem-Like Cell Line Derived from a Non-hematopoietic Population of Humane Ubilical Cord Blood. *Stem Cell and Development* 15, 391–406 (2006)
6. Castleman, K.: *Digital Image Processing*. Prentice Hall, Englewood Cliffs (1996)
7. Comaniciu, D., Meer, P.: Cell image segmentation for diagnostic pathology. In: Suri, J.S., Setarehdan, S.K., Singh, S. (eds.) *Advanced algorithmic approaches to medical image segmentation: state-of-the-art application in cardiology, neurology, mammography and pathology*, pp. 541–558 (2001)
8. Du, Q., Faber, V., Gunzburger, M.: Centroidal Voronoi tessellations: Applications and algorithms. *SIAM Rev* 41, 637–676 (1999)
9. Duda, R.O., Hart, P.E.: *Pattern Classification and Scene Analysis*. John Wiley and Sons, New-York (1973)
10. El-Sakka Mahmoud, R., Kamel Mohamed, S.: Adaptive Image Compression Based on Segmentation and Block Classification. *Int. Journal of Imaging Systems and Technology* 10(1), 33–46 (1999)
11. Garbay, C., Chassery, J.M., Brugal, G.: An interactive region growing process for cell image segmentation based on local color similarity and global shape criteria. *Anal. Quantit. Cytol. Histol.* 8, 25–34 (1986)
12. Hartigan, J.: *Clustering Algorithms*. Wiley Interscience, New York (1975)
13. Hartigan, J., Wong, M.: Algorithm AS 136: A k-means clustering algorithm. *Appl. Stat.* 28, 100–108 (1979)
14. Inaba, M., Katoh, N., Imai, H.: Applications of weighted Voronoi diagrams and randomization to variance-based k-clustering. In: *Proc. Tenth Ann. Symp. on Computational Geometry*, pp. 332–339 (1994)
15. Jain, A., Dubes, R.: *Algorithms for Clustering Data*. Prentice Hall, Englewood Cliffs (1988)
16. Jain, A., Murty, M., Flynn, P.: Data Clustering: A Review. *ACM Computing Surveys* 31(3), 264–323 (1999)
17. Jiang, K., Liao, Q.M., Dai, S.Y.: A novel white blood cell segmentation scheme using scale-space ltering and watershed clustering. In: *Proc. Int. Conf. on Machine Learning and Cybernetics*, vol. 5, pp. 2820–2825 (2003)
18. Liao, Q., Deng, Y.: An accurate segmentation method for white blood cell images. In: *Proc. Int. Symposium on Biomedical Imaging*, pp. 245–248 (2002)
19. Liedtke, C.E., Gahm, T., Kappei, F., Aeikens, B.: Segmentation of microscopic cell scenes. *Analyt. Quant. Cytol. Histol.* 9, 197–211 (1987)
20. Lockett, S.J., Herman, B.: Automatic detection of clustered, fluorescent-stained nuclei by digital image-based cytometry. *Cytometry* 17, 1–12 (1994)

21. Malpica, N., Ortiz, C., Vaquero, J.J., Santos, A., Vallcorba, I., García-Sagredo, J.M., Pozo, F.: Applying watershed algorithms to the segmentation of clustered nuclei. *Cytometry* 28, 289–297 (1997)
22. Nilsson, B., Heyden, A.: Model-based segmentation of leukocyte clusters. *Proc. Int. Conf. on Pattern Recognition* 1, 727–730 (2002)
23. Okabe, A., Boots, B., Sugihara, K.: *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Wiley, Chichester (1992)
24. Ongun, G., Halici, U., Leblebicioglu, K., Atalay, V., Beksac, M., Beksac, S.: An automated differential blood count system. In: *Proc. Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, vol. 3, pp. 2583–2586 (2001)
25. Phillips, S.J.: *Acceleration of k-means and Related Clustering Algorithms*. LNCS. Springer, Heidelberg (2002)
26. Proffitt, R.T., Tran, J.V., Reynolds, C.P.: A fluorescence digital image microscopy system for quantifying relative cell numbers in tissue culture plates. *Cytometry* 24, 204–213 (1996)
27. Rasmussen, E.: *Clustering Algorithms*. In: Frakes, W.B., Baeza-Yates, R. (eds.) *Information Retrieval: Data Structures and Algorithms*, Prentice Hall, Englewood Cliffs (1992)
28. Zadeh, L.A.: Fuzzy Sets. *Inform. Control* 8, 338–353 (1965)
29. Zhang, Y.J.: A Survey on Evaluating Methods for Image Segmentation. *Pattern Recognition* 29(8), 1246–1335 (1996)

Application to Estimate Haplotypes for Multiallelic Present-Absent Loci

Robert Nowak

Electronics Systems Institute, Warsaw University of Technology
Nowowiejska 15/19, 00-665 Warsaw
r.m.nowak@elka.pw.edu.pl

Summary. The article presents an algorithm and an application to estimate haplotype frequencies from genotype data for unrelated individuals. Presented approach can handle loci with multiple alleles as well as silent (null) alleles. The mathematical model and an expanded Expectation-Maximization algorithm is described. The computer program, called NullHap, available freely at <http://staff.elka.pw.edu.pl/~rnnowak2/nullhap> implements presented ideas. Comparison with known software to estimate haplotypes: Arlequin, PHASE and Haplo-IHP proves the advantage presented method.

1 Introduction

Laboratory techniques used to determine local haplotypes [4] are often too expensive for large-scale studies. The lack of phase information (as depicted in Fig. 1) provided by the popular typing methods could be overcome by using likelihood-based calculations [3], which estimates haplotype frequencies in population, and reconstruct of haplotype pair in an individual. This approach is more cost-effective and powerful than linkage analysis [2], and gives more information than single marker-based methods [7].

The maximum likelihood approach to haplotype estimation from unrelated individuals was widely discussed, and a number of programs is available for that

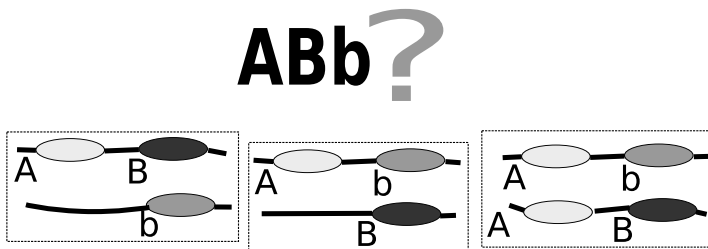


Fig. 1. The lack of phase information and ambiguous of the heterozygous status for present-absent genes after popular typing techniques

purpose. The Expectation - Maximisation algorithm (EM), proposed in [6] is the most popular (used in tested applications Arlequin [5] and PHASE [9]), but also others methods were applied: Bayesian networks [8], Monte Carlo [1] and Hidden Markov Model.

The polymorphism for present-absent genes, where ambiguous as regards to the heterozygous status of the individuals (Fig. 1) is poorly analyzed in literature. The only available computer program designed to handle this problem, proposed in [10] can estimate haplotypes only for biallelic loci.

2 Algorithm

In a diploid organism, when the k polymorphic loci is analyzed and each locus has l_i different variants (optionally including null variant) for i -th locus, the number of different haplotypes is $H = \prod_{i=1}^k l_i$, and the number of different haplotype resolutions is $R = \frac{1}{2}H * (H + 1)$. The lack of phase information as well as ambiguous to the heterozygous state for locus with null variant cause that only the G (equation 1) different genotypes could be observed.

$$G = \prod_{i=1}^k \frac{(l_i - \delta_i)(l_i + 1 - \delta_i) + 2\delta_i}{2}, \delta_i = \begin{cases} 1 & \text{locus with null allele} \\ 0 & \text{locus without null allele} \end{cases} \quad (1)$$

The number of haplotype resolutions r_j which gives genotype j (when phase information is lost) is described by equation 2, where s_j is the number of observed heterozygous and t_j is the number of observed (not null) alleles for loci with null allele. The number of haplotype resolutions grows exponentially with number of observed loci, thus full space search can not be applied.

$$r_j = \begin{cases} 2^{s_j-1} * 3^{t_j} & \text{for } s_j > 0 \\ \frac{3^{t_j+1}}{2} & \text{for } s_j = 0 \end{cases} \quad (2)$$

2.1 Maximum Likelihood Approach to Estimate Haplotypes

A sample of genotype data from unrelated n individuals is simplified to a vector $S = (n_1, n_2, \dots, n_G)$, where G is the number of different genotype data (with lack of phase information), and n_j is the number of individuals having j -th genotype, $\sum_{j=0}^G n_j = n$. In maximum likelihood approach haplotype probability h_i is estimated to maximise the probability of the given sample (equation 6).

The condition probability of sample S , given each genotype probability g_i , assume unrelatedness of individuals in sample is provided in equation (3), where α not depends on g_i .

$$P(S | g_1, g_2, \dots, g_G) = \frac{n!}{n_1! * n_2! * \dots * n_G!} * \prod_{j=1}^G g_j^{n_j} = \alpha \prod_{j=1}^F g_j^{n_j} \quad (3)$$

The probability of genotype g_j is the sum of probabilities of respective haplotype pairs z_{mn} , and with Hardy-Weinberg assumption, is calculated from haplotype probabilities, as shown in equation 4, where z_{mn} is the probability of haplotype pair m and n , r_j is a number of haplotype pairs for given j -th genotype, and h_m, h_n are the probabilities of haplotypes m and n respectively.

$$g_j = \sum_{i=0}^{r_j} z_{mn}, \text{ where } z_{mn} = \begin{cases} h_m^2 & \text{for } m = n \\ 2 h_m h_n & \text{for } m \neq n \end{cases} \quad (4)$$

The estimation of haplotypes probability to maximise the probability of observed sample can be described as optimization, the equation 6 summarizes considered approach.

$$\begin{aligned} \arg \max_{h_1, h_2, \dots, h_H} P(S | h_1, h_2, \dots, h_H) &= \arg \max_{h_1, h_2, \dots, h_H} \prod_{j=1}^G \left(\sum_{i=0}^{r_j} z_{mn} \right)^{n_j}, \\ z_{mn} &= \begin{cases} h_m^2 & \text{for } m = n \\ 2 h_m h_n & \text{for } m \neq n \end{cases} \end{aligned} \quad (5)$$

2.2 Extended EM Algorithm

An Expectation-Maximization algorithm alternates between performing an expectation step $E^{(t)}$, which computes an expectation value of unknown parameters, here the probabilities of haplotype pairs, and a maximization step $M^{(t)}$, which computes the value of parameters by maximising the probability of observed data. The parameters found on the $M^{(t)}$ step are then used to begin another $E^{(t+1)}$ step, and the process is repeated until the parameters are changed.

The EM algorithm peaks at local optimum. In presented solution this algorithm is used to maximize equation 6. The algorithm details for polymorphism with k observed loci, l_i variants for i -th locus, and sample $S = (n_1, n_2, \dots, n_G)$ are supplied below.

Initiation

The initial haplotype pair probabilities (the E^0 step) are set as described in equation 6. For each haplotype pair the probability depends only on the number of resolutions for given genotype. It is similar to initiation in [6].

$$z_{mn}^{(0)} = \frac{1}{r_j} \text{ where the } mn \text{ gives the } j \text{ genotype} \quad (6)$$

Maximization step

The genotype probability is calculated as sum of responding haplotype pairs probabilities, and the next values of haplotype pair probabilities are calculated to maximize given sample probability. Details in equation 7.

$$z_{mn}^{(t+1)} = \frac{n_j}{n} * \frac{z_{mn}^{(t)}}{g_j^{(t)}}, \text{ where } mn \text{ gives genotype } j, g_j^{(t)} = \sum_x^{r_j} z_x^{(t)} \quad (7)$$

Expectation step

Haplotype probabilities h_m are calculated from given haplotype pair probabilities z_{mn} , it is a half of sum of each haplotype pair probability when given haplotype occurs. The next expected haplotype pair probabilities are calculated using haplotype probabilities as described in equation 8.

$$z_{mn}^{(t+1)} = \begin{cases} (h_m^{(t)})^2 & \text{for } m = n \\ 2 h_m^{(t)} h_n^{(t)} & \text{for } m \neq n \end{cases} \text{ where } h_m^{(t)} = \frac{1}{2} (\sum_i z_{im}^{(t)} + \sum_j z_{mj}^{(t)}) \quad (8)$$

Stop conditions

Algorithm stops when the stability of estimations between following steps are obtained. In presented approach it means the absolute difference between calculated probabilities is less then ϵ (equation 9).

$$\sum_{i=1}^R |z_i^{(t+1)} - z_i^{(t)}| < \epsilon \quad (9)$$

Finally, the haplotype probabilities are calculated (by another E step), as well as the conditional probability of the haplotype pair, given genotype are estimated using equation 10.

$$z_{mn}|g_j = \frac{z_{mn}}{g_j} = \frac{z_{mn}}{\sum_x^{r_j} z_x} \quad (10)$$

3 NullHap: Validation and Comparison with Other Applications to Haplotype Estimation

Considered algorithm was used in application called NullHap. The implementation in C++, depends on the Boost libraries and faif library [12]. The source code and binaries for Windows NT/2000/XP and Debian Linux are available at project web page [11].

Application was tested on simulated and real data sets. Results were compared with previously mentioned programs: Arlequin [5], PHASE [9] and Haplo-IHP [10].

The main advantage of NullHap over others known applications is ability to handle problems when multiallelic locus containing null variant (Tab. 1).

Test on Generated Data Sets

To compare the ability to calculate the haplotypes probabilities for considered applications the most probable sample was generated for five polymorphisms

Table 1. Short comparison of analyzed applications to estimate haplotypes

name	biallelic multiallelic null variants		
Arlequin	+	+	-
PHASE	+	+	-
Haplo-IHP	+	-	+
NullHap	+	+	+

Table 2. Example of simulated data. Two loci polymorphism was considered: probabilities were assumed for each haplotype, next the sample was generated, than haplotype probabilities were estimated, finally results were compared.

haplotype	probability h_i				
	assumed	Arlequin	PHASE	Haplo-IHP	NullHap
A0B0	0.2	0.042	0.04	0.25	0.2
A0B1	0.2	0.126	0.12	0.25	0.2
A1B0	0.2	0.147	0.14	0.25	0.2
A1B1	0.2	0.442	0.42	0.25	0.2
A2B0	0.1	0.021	0.07	0	0.1
A2B1	0.1	0.221	0.21	0	0.1
error	-	77%	67%	50%	0%

Table 3. Haplotype estimation probability error for considered applications

No	example description	error			
		Arlequin	PHASE	Haplo-IHP	NullHap
1	biallelic loci: A(A0, A1), B(B0, B1), C(C0, C1), null variants: A0, B0 and C0	61%	50%	0.7%	0%
2	biallelic loci: A(A1, A2), B(B1, B2), no null variants	0%	0%	0%	0%
3	multiallelic loci: A(A0, A1, A2), B(B0, B1, B2), null variants: A0, B0	62%	62%	100%	0%
4	multiallelic loci: A(A1, A2, A3), B(B1, B2, B3), no null variants	0%	1%	78%	0%
5	multiallelic loci, A(A0, A1, A2), B(B0, B1), null variants: A0, B0	77%	67%	50%	0%

with varying locus characteristics. An example of assumed and estimated probabilities is shown in Tab. 2 (others examples are available at web [11]). In Tab. 3 the results of simulations are summarized. The error was calculated as provided in equation 11, where x is assumed probability, and x^* is calculated one.

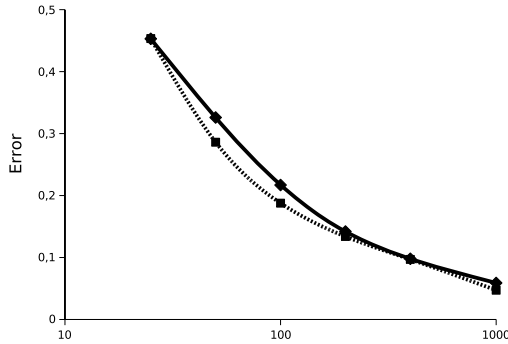


Fig. 2. Sample size effect on organism described in tab. 2. The figure describes the median of error in function of sample size for two examples (example 4 and 5 from tab. 3).

Table 4. The haplotype estimations for HLA data set for 99 Polish [5]. The difference between estimated probabilities between programs is less then 2%.

haplotype	Arlequin	NullHap	Phase	haplotype	Arlequin	NullHap	Phase
1500 0602	0.16	0.156	0.153	1301 0603	0.055	0.055	0.05
0700 0200	0.1	0.101	0.101	0401 0302	0.035	0.035	0.035
0101 0501	0.091	0.091	0.091	0104 0501	0.035	0.035	0.035
0301 0200	0.076	0.076	0.075	0806 0402	0.025	0.025	0.025
1101 0301	0.066	0.066	0.063	0700 0303	0.02	0.02	0.02

$$error = \frac{1}{N} \sum_{i=1}^N \left| \frac{x - x^*}{x} \right| \tag{11}$$

Next, the effect of sample size on the exactness of the method were investigated, therefore the minimal sample size required to provide reliable estimations could be obtained. The 10 samples of 25,50,100,200,400 and 1000 from an infinite population in Hardy - Weinberg equilibrium characterized by haplotype distribution as those given 4-th and 5-th example in tab. 3. The haplotype probability were estimated, and median of errors (equation 11) are illustrated in fig. 2.

Test On Real Data Sets

The applications able to analyze multiallelic loci (Arlequin, PHASE and NullHap) were used to calculate HLA (DRB1-DQB1) data set for 99 Polish supplied by [5]. The case contains two multiallelic loci (36 and 14 variants), without null variants. The results, shown in tab. 4, are very similar for each computer program.

Table 5. The haplotype estimations for KIR data set, 200 Irish, [10]. The difference of estimated probabilities between programs is about 3%.

2DS2	2DL2	3DL3	2DL5B	2DL1	3DS1	Haplo-IHP	NullHap
0	0	1	0	1	0	0.546	0.55
0	0	1	0	1	1	0.101	0.1
1	1	0	1	1	1	0.095	0.094
1	1	0	0	0	0	0.075	0.083
1	1	0	0	1	0	0.064	0.056
1	1	0	1	1	0	0.028	0.028
0	0	0	0	1	0	0.019	0.028
1	1	1	1	1	0	0.021	0.021
1	1	1	0	1	1	0.019	0.019
0	0	0	1	1	1	0.009	0.01

The data set containing the KIR loci for 200 Irish [10], was used to check the applications takes into account null variants (Haplo-IHP and NullHap). Each locus is biallelic and has null variants. Results provided in tab. 5 shows similar results.

Tests of simulated samples proves ability to estimate assumed haplotype probabilities. In case of polymorphisms able to calculate by other known computer programs the results produced by NullHap were similar.

Performance Tests

The analysis of computational time in different scenarios is presented in tab. 6. All computations were achieved on Celeron M 1.6GHz, 1GB RAM, under Debian Linux or Windows XP. The number of loci and sample size influence to computational time.

Table 6. Computational time for considered applications (HaploIHP in parenthesis with greedy algorithm). Results only for applications able to estimate haplotypes for given polymorphism.

number of			time for application			
loci	haplotypes	observations	Arlequin	Phase	HaploIHP	NullHap
2	6	100	0.13s	46s	0.5s	0.07s
2	9	100	0.06s	47s	0.15s	0.04s
3	8	100	0.04s	69s	0.58s	0.02s
2	504	99	0.22s	53s	-	37s
2	540	99	0.34s	58s	-	39s
7	128	200	-	-	145s	13s
9	512	200	-	-	1300s(8s)	209s
11	2048	200	-	-	24h(10s)	3h

4 Conclusion

Presented application can effectively estimate haplotype probabilities with performance similar to other computer programs. It should be emphasized that the main advantage of created application is the ability to estimate haplotypes for every type of polymorphism, in particular polymorphisms with multiallelic loci with null variants. The tests for simulated data shows that NullHap obtain results closer to assumed than others computer programs.

The demonstrated application is under development, and some improvements are planned such as the ability to consider the cases of missing data and the additional step removing unimportant haplotypes, to speed-up computations. The new versions will be available at project web side.

The application would be confirmed by larger studies involving experimental work on allele typing.

References

1. Boettcher, P., Pagnacco, G., Stella, A.: A Monte Carlo Approach for Estimation of Haplotype Probabilities in Half-Sib Families. *J. Dairy Sci.* 87, 4303–4310 (2004)
2. Botstein, D., Risch, N.: Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. *Nat.Genet.* 33, 228–237 (2003)
3. Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society* 39, 1–39 (1977)
4. Douglas, J., et al.: Experimentally derived haplotypes substantially increase the efficiency of linkage disequilibrium studies. *Nat. Genet.* 28, 361–364 (2001)
5. Excoffier, L., Laval, G., Schneider, S.: Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1, 47–50 (2005)
6. Excoffier, L.M., Slatkin, M.: Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.* 12, 921–927 (1995)
7. Morris, R., Kaplan, N.: On the advantage of haplotype analysis in the presence of multiple disease susceptibility alleles. *Gen. Epidemiol.* 23, 221–233 (2002)
8. Niu, T., Qin, Z., Xu, X., Liu, J.: Bayesian haplotype inference for multiple linked single nucleotide polymorphisms. *Am. J. Hum. Genet.* 70, 157–169 (2002)
9. Stephens, M., Smith, N., Donnelly, P.: A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68, 978–989 (2001)
10. Yoo, Y., Tang, J., Kaslow, R., Zhang, K.: Haplotype inference for present-absent genotype data using previously identified haplotypes and haplotype patterns. *Bioinformatics* 23, 2399–2406 (2007)
11. NullHap, <http://staff.elka.pw.edu.pl/~rnowak2/nullhap/>
12. The faif library, <http://faif.sourceforge.net/>

Detection of Mitotic Cell Fraction in Neural Stem Cells in Cultures

Anna Korzynska¹ and Marcin Iwanowski²

¹ Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, ul. Ksiecica Trojdena 4, 02-109 Warszawa, Poland

Anna.Korzynska@ibib.waw.pl

² Institute of Control and Industrial Electronics, Warsaw University of Technology, ul. Koszykowa 75, 00-662 Warszawa, Poland

iwanowski@isep.pw.edu.pl

Summary. Automation of monitoring and analysis of cell culture condition is crucial for fast and reliable optimization of culturing methods e.g. in the regenerative medicine. So the method of automatic cells counting during culture monitoring and analysis of cells population, according to the model of cell culture development, is needed. To solve this general problem several separated problems should be investigated. One of them is the detection of cells which are soon going to divide, the moment when in the new frame two daughter cells could be expected in the place occupied by a single cell in the previous image. A method of localization of cells, which potentially would divide, is proposed in the paper. The mathematical morphology operations are used to detect small, rounded and converged cells which *in potentia* could be mitotic in a single frame from the sequence of images. The proposed method is applied to images of neural stem cells, made with the laser scanner confocal microscope. Results of the experiments are also presented in the paper.

1 Introduction

Since stem cells are a potential source of cells for use in the regenerative medicine, the monitoring and analysis of the culture quantity is crucial for reliable optimization of culturing methods [4]. Monitoring of the living cell cultures by means of a human-made evaluation at the microscope is typical method used in the laboratory as the fastest and cost free [1, 22]. The acquired information is supported temporary by immunohistochemistry and flowcytometry culture evaluation. The idea of the continuous culture condition monitoring using a computer supported microscopy seems is promising because it allows evaluating the culture growth and to discriminate cells revival or differentiation in the observed culture as early as possible. The automation of the cells culture observation and evaluation is the main goal which could be achieved by solving several separated problems. One of them is to investigate a method of automatic cells counting and in microscopic images acquired in the constant time increments. Next, based on this information, the cell culture growth dynamics, in a sense of the rates of cells proliferation and death or so called mitotic index (defined as a quotient of the

number of mitotic and normal cell) [1], could be evaluated. To do that, one needs to identify all cases of cell divisions in cells population observed in the image sequence.

Because of some inherent properties of both the process of collection data in time sequence with a constant time increments and the process of cell division itself (described in the next section) it is impossible to detect each cell division analyzing each single frame. The moment of the daughter cells separation could be missing in the observation gaps. So the idea of detection of cells, which would have divided in the nearest future allows to monitor both the number of cell divisions and the trees of cells linkage [1, 4].

The goal of investigations described in this paper is to detect and to localize all cells which would have divided in the nearest future in the image plane based on the features of cell shape, area and texture. The proposed method makes use of the advanced morphological image processing tools [16] and consists of three steps. First, the nonuniform background, which is characteristic feature of most of the images, is removed using morphological top-hat operator. In the second step the image is simplified using morphological tools based on reconstruction operator in order to make cell region more homogeneous. Finally, image is binarized using double threshold technique.

The paper is organized as follows. Sect. 2 contains the literature review, Sect. 3 described the biological background. In the Sect. 4 the morphological image processing tools are described and the proposed method is introduced. Sect. 5 describes the experiments and finally, Sect. 6 concludes the paper.

2 Related Works

There are some applications which offer mitotic cell selections, manual methods [6] as well as the semiautomatic ones based on image processing [12]. Both of them can be useful only if any off line cells counting is possible. In the case of continuous monitoring a fully automatic method is required.

Automatic methods for particle detection used in the cryo-electron microscopy, which seems to be similar to the problem of cell localization, have been investigated for a long time. There are descriptions of methods based on intensity or texture comparison, on cross-correlation of template matching [20] or neural networks in Nicholson and other reviews [14]. The Authors come to a conclusion that none of presented methods alone is good enough. Consequently, their successors have used the hybrid method [3, 9], supported by extra information of the local average intensity method [8], cross-correlation with pruning the list of candidates by examining modified local gradient [13] or agents based systems [2] in which agents performing their action locally are used.

For a long time, the problem of microscopic images of living cells segmentations has been solved by detection of edges or regions [1]. Nowadays there are two directions of developing progress in this field: (1) using a priori knowledge and models [13], (2) employing the cooperation between methods, and their adaptation to the local situations in the image [5]. Within the second direction there

are two methods going back to the edge and region approach: combined texture and edge based method, proposed by one of the authors and her collaborators [9] and cooperative system [1].

The direct approach uses the cell segmentation and detection according to the cell features, most important of which are the following:

- shape [1, 6, 8, 10]: enumeration and detection of proliferating bone marrow progenitors, cancer cells, neutrophils or artefact in smears;
- texture [9, 10]: identify nuclei by characterizing the chromatin structure;
- color [8, 15]: nuclei in fluorescent-stained samples.

3 Microscopic Investigation of the Cell Growth

3.1 Cell Growing Process

According to the knowledge about cells divisions [1, 22], it is a long-lasting process. It starts from changes in the cell inner reorganization (prophase and metaphase) which become visible as morphological continuous changes in the cell shape, area and texture. The cell converges and rounds shape up to become a sphere (anaphase). Next, it narrows in the middle and becomes eight shaped with gradually more and more narrow waist/neck (telophase), which finally brakes what causes that two separate converged cells appear. Characteristics of cells division during the stage of eight shaped cell could not be visible in the image sequence acquired in constant time increments. To find all cases of cell divisions, the converged cells should be detected as a cell which can divide *in potentia*. Because a cell can be converged not only just before its division, so behavior of the rounded converged cell in the next few frames should give us an answer if the cell has divided or has not. Because of the physiology of the cell division these cells differ in appearance from most but not all not dividing cells. They are small, rounded, converged with specific texture and strong halo while most of other stem cells adhere to the dishes and start to develop pseudopodia, to elongate in various degree up to become flattened (see Fig. 1).

The cell division is denoted if two small, rounded, converged cells occupy a place which, in the previous frame, has been occupied by a cell which potentially could divide. It is observed that stem cells in culture *in vitro* sometimes aggregate in clusters, in which they are located closely one to another, but sometimes migrate out of clusters, see all figures in this paper [24]. The detection of each particular cell is much more difficult if it is located in the cluster (shown in Fig. 1 and Fig. 3) than if the cell is surrounded by the background (shown in Fig. 2). So that if converged cells are detected in cells cluster, the additional steps of cells separation, cell shape and area measurements are needed.

3.2 Characteristics of the Images Sequence and Cells in the Images

Two types of microscopic techniques are used in observation of living cells: the conventional light microscopy (bright field, phase contrast, dark field or Nomarski contrast microscopies) or the confocal microscopy with coherent laser

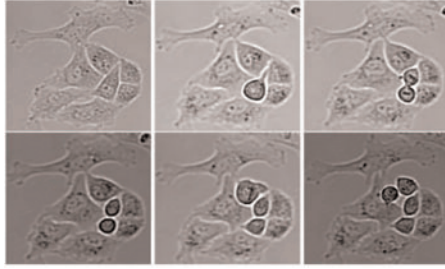


Fig. 1. Six images fragments from the images sequences acquired with red laser confocal microscopy show both the variation in image quality, lighting conditions, focus plane during 4 hours 15 minutes observations and the divisions of two cells located in the cells cluster

light [5, 10]. Both techniques insignificantly influence the cells behavior. The cells make transparent objects so they have a poor contrast with respect to the gray background in both microscopic techniques: the laser scanning confocal and the conventional brightfield microscopies. Furthermore, the quality of recorded images depends on the illumination at present, on the internal inhomogeneous light distribution in the microscopic light pathway and on the focus plane position; all of them are instable during the long-term observation. To observe all of these features of images, see Fig. 1, where fragments of images from the sequence acquired with the red laser confocal microscopy [Zeiss] are presented.

The fluctuation in intensity of the gray level is observed within cells regions as well as within the background region. But variations of the gray level within cell regions have a larger amplitude than in the background. It is because the variations due to noise and light distribution, observed in the background, are dumped by various light transmission through the cellular organelles in the cell body regions.

The cells population observed in images is not homogeneous, cells differ in size, shapes and texture properties. The investigations are concentrated on the specific fraction of cells - small, rounded and converged ones, because they could divide in the nearest future. Cells with different morphology e.g. flattened cells (Fig. 1. the cell on the top) and cells which start to develop pseudopodia and to elongate but which are still far from being flattened (Fig. 1. right side cells) should be excluded from consideration even, if some of their part resembles converged cells in the texture properties.

The main features which discriminate a cell under investigation from other cells are size, shape, converged cellular body remarkably darker than the background and with irregularly but closely and densely dispersed bright and dark dotes and with the strongest halo around them. The last two are principally exploited in the proposed method.

4 Proposed Method

4.1 Mathematical Morphology Methods in Image Processing

Image processing by using mathematical morphology [15, 17, 18, 19, 21, 22] is based on two nonlinear local operators: minimum (infinimum) and maximum (supremum) applied within the neighborhood of every image pixel. This neighborhood is defined using the structuring element. The popular structuring elements covers pixel's neighborhood of a given radius. In digital grid various ways of defining such a neighborhood are considered. Two principal morphological operators, dilation and erosion of image f with structuring element B are defined as, respectively:

$$\delta_B(f)[p] = \max_{b \in B} [f(p + b)] ; \varepsilon_B(f)[p] = \min_{b \in B} [f(p + b)], \quad (1)$$

where p stands for the single image pixel. The morphological filters of opening and closing contain the combination of morphological dilation and erosion performed one after another:

$$\gamma_B(f) = \delta_{B^T}(\varepsilon_B(f)) ; \varphi_B(f) = \varepsilon_{B^T}(\delta_B(f)), \quad (2)$$

where γ represents the operator of opening and φ - the operator of closing, B^T stands for transposed structuring element B . Opening and closing remove from the image its elements (object, noise) lighter and darker then the background, respectively. By combining opening and closing the alternating filters are defined. They are applied to remove both darker and lighter elements of the image.

Opening and closing are also used to detect objects while removing the image background by means of the two kinds of top-hat operators. defined as follows:

$$WTH_B(f) = f - \gamma_B(f) ; BTH_B(f) = \varphi_B(f) - f, \quad (3)$$

where WTH stands for white top-hat and BTH - for the black top-hat. The description 'white' ('black') refers to type of objects that are detected by a particular operator - lighter (darker) than the background. In order to extract from the background both darker and lighter objects, a sum of results of both top-hats can be used:

$$TH_{B,B'}(f) = WTH_B(f) + BTH_{B'}(f) = \varphi_{B'}(f) - \delta_B(f). \quad (4)$$

As shown above, sum of both top-hats is simply a difference between closing and opening.

The effect of filtering by opening and closing implies an important disadvantage of shapes modification. To avoid this effect, advanced morphological filters based on the morphological reconstruction are used. To define a morphological reconstruction one has to introduce the geodesic erosion and dilation. Contrary to the classic erosion and dilation (Eq.1), the geodesic ones require two input images. The second of them - called a mask - restricts the area in which the

operator is performed. The geodesic dilation of size 1 (resp. geodesic erosion of size 1) is defined as:

$$\delta_{B,g}(f) = \delta_b(f) \vee g ; \varepsilon_{B,g}(f) = \varepsilon_B(f) \wedge g. \quad (5)$$

The \wedge and \vee operators stand for the point-wise minimum and maximum. The geodesic erosion and dilation of a given size n are defined as, respectively:

$$\varepsilon_{B,g}^{(n)}(f) = \underbrace{\varepsilon_{B,g}(\varepsilon_{B,g}(\dots\varepsilon_{B,g}(f)\dots))}_{k\text{-times}} ; \delta_{B,g}^{(n)}(f) = \underbrace{\delta_{B,g}(\delta_{B,g}\dots\delta_{B,g}(f)\dots)}_{k\text{-times}} \quad (6)$$

The image created by the successive erosions/dilations has an important property - for certain n the resulting image stops to change - the idempotence is reached. Owing to that feature, the reconstruction operator can be defined. The reconstruction by dilation is thus defined as:

$$R_{B,g}^\delta(f) = \delta_{B,g}^{(i)}(f) ; i = \operatorname{argmin}_j \left\{ \delta_{B,g}^{(j)}(f) = \delta_{B,g}^{(j+1)}(f) \right\} \quad (7)$$

By duality, the reconstruction by erosion is defined as:

$$R_{B,g}^\varepsilon(f) = \varepsilon_{B,g}^{(i)}(f) ; i = \operatorname{argmin}_j \left\{ \varepsilon_{B,g}^{(j)}(f) = \varepsilon_{B,g}^{(j+1)}(f) \right\} \quad (8)$$

Image f in both above operators is usually called a marker image. Structuring element B used in geodesic operators: geodesic erosion and dilation as well as in the reconstruction is usually defined as the closest 4- or 8- connected pixel's neighborhood (elementary structuring element in 4- or 8-connected grid).

By means of reconstruction operators, the filters of opening by reconstruction and closing by reconstruction are defined as, respectively:

$$\gamma_{B,B'}^{rec}(f) = R_{B',f}^\delta(\varepsilon_B(f)) ; \varphi_{B,B'}^{rec}(f) = R_{B',f}^\varepsilon(\delta_B(f)) \quad (9)$$

Their filtering results are similar to those of simple opening and closing defined by the Eq.3, but without the effect of shape modification, which is their great advantage. Both filters can be used instead of classic ones in top-hat operators defined by Eq.4 and Eq.5. In the latter case appropriate top-hat will be called farther in this text: *WTHR*, *BTHR* and *THR* for white and black top-hats and sum of top-hats by reconstruction respectively.

The second application of reconstruction which is used in the proposed method is conversion of graytone image into a binary one. The simple conversion is performed using the thresholding operator - pixels of grayvalues higher than given threshold are set to 1, all the other - to 0:

$$T_\alpha(f)[p] = \begin{cases} 1 & \text{if } a \geq f(p) \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

This approach has several disadvantages, which are important for particular kind of images. One of them is following the dependence between unnecessary

areas on the binary image and the shape of the correctly binarized regions. For lower threshold all the required areas are visible, but their shape is not correct. On the other hand, for higher threshold the shape of required areas is correct but there exist also some unnecessary and unwanted areas on the binary image. To solve that problem, the double threshold method is used. This operation consists of two simple thresholdings with two threshold levels (narrow and wide) followed by the reconstruction of the binary image with narrow thresholds and with a mask equal to the binary image constructed with wide one. It is than equal to:

$$DT_{[a,b]}(f) = R_{T_b}^{\delta}(T_a(f)) \tag{11}$$

where $a \leq b$ are threshold values.

4.2 Proposed Method Workflow

The proposed method makes use of the morphological tool described in previous section. The whole algorithm can be functionally divided into 3 steps: background suppression, image simplification, binarization. They are described below.

Background Suppression

The input images are characterized by non-uniform background. The variations of the graytones of the background area are however much smoother than changes of graytones related to the presence of cells. In order to perform further with the detection, the background must be thus removed, which is performed by means of the top-hat operator. Another two features of the initial image determine the type of top-hat to use. First of them is the fact that investigated cells are characterized by both the inner part darker and the halo-effect which is lighter than the background. This observation makes the choice of the sum of top-hats (Eq.5) obvious in this case. The second important issue is the fact that the background removal cannot influence the shapes of cells. Usage of classical top-hats based on openings and closings obviously would modify the cells' shape, which could influence the shape of cells detected in the segmentation phase. In order to avoid this problem, the openings and closings by reconstruction are used instead:

$$f_1 = BTHR_{B1, BR}(f) + WTHR_{B2, BR}(f) = \varphi_{B1, BR}^{rec}(f) - \gamma_{B2, BR}^{rec}(f) \tag{12}$$

Where f stand for the input image (shown in Fig. 2(a)), f_1 for the result of above processing (Fig. 2(b)). $B1$ and $B2$ represent structuring elements used in erosion/dilation. BR is the structuring element used in reconstruction. The choice of these and other parameters as well as the choice of structuring elements will be discussed later, in the next section.

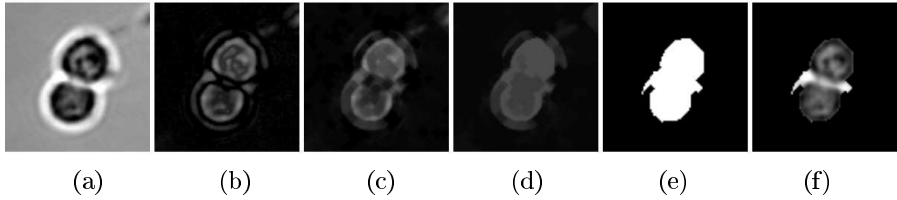


Fig. 2. The sequence of image operations in proposed method (description in the text)

The resulting image without background contains the cells with clearly shaped borders what is presented in Fig. 2(b). In case of cells having the halo-effect, the boundary is split into two light outlines separated by a narrow dark gap. This dark inclusion can be a problem for further processing, so closing operator of a small size is used to remove it:

$$f_2 = \varphi_{B3}(f_2). \quad (13)$$

The result of this operation is presented in Fig. 2(c).

Simplification

Although the image obtained in the preceding step does not contain the background, it still contains a noise - small regions which are not cells, which would strongly influence the segmentation process resulting in oversegmentation. In order to remove unwanted details from the image, the alternating sequential filter is applied consisting of openings and closings by reconstruction. The application of reconstruction filters guarantees that the shape of cells will not be deformed. The following equation describes applied filter:

$$f_3 = \varphi_{B5,BR}^{rec}(\gamma_{B5,BR}^{rec}(\varphi_{B4,BR}^{rec}(\gamma_{B4,BR}^{rec}(f_2)))). \quad (14)$$

Apart from removal of unwanted objects, smaller than cells, the filtering step modifies the inner part of cells. It is presented in Fig. 2(d). Owing to this modification, the grayvalues of pixels inside cells depend on the intensity of texture within the same area on the input image. The more dense and contrasted the texture is on the input image, the higher grayvalue characterizes this area on f_4 image. Thus, the level of 'convergence' (or 'flatness') of a cell is mapped into average grayvalue of the appropriate region of the simplified image.

Binarization

Due to, above mentioned, properties of f_3 image it is perfectly suited to segmentation by thresholding. Simple thresholding according to the Eq.10 does not produce correct effects. The reason is non fully homogeneous graylevels inside cells. The inner parts of cells are usually lighter, while those closer to the cell

boundary - darker. On the other hand some less converged cells are characterized by darker grayvalues. Owing to that, one can either set lower threshold and obtain correctly segmented converged cells and - in addition - some transitional, or get only converged cells (without transitional) but of inappropriate shape - too small. This is typical situation in which the double thresholding can be applied. The thresholding according to the Eq.11 is applied with thresholds set as parameters.

$$f_4 = DT_{[a,b]}(f_3), \tag{15}$$

where a and b stands for thresholds: lower and upper. The result of this operation is presented in Fig. 2(e) while in Fig. 2(f) the result of threshold superimposed on original image is presented.

5 Experiments

5.1 Experimental Material

The 5 image sequences of neural stem cells culture were collected using an inverted confocal laser scanner microscope with red color equipped with a temperature, CO_2 and humidity controlled incubation chamber with perfusion [Zeiss]. Cells were seeded in density 1000 cell/ml few hours before observation and were stayed in standard conditions (37°C, 5% CO_2 and humidity 95%) before and during the time of observation. Images of 2048x2048 pixels were acquired in every 15 or 20 minutes up to 46 hours. They document chosen observation plane which covers only 120x120 μm space on culture dish. 6 images was randomly chosen to perform the evaluation of the method. The evaluation was done on the decreased resolution images of 1024x1024 pixels (bicubic smoother resampling).

5.2 Parameters Choice

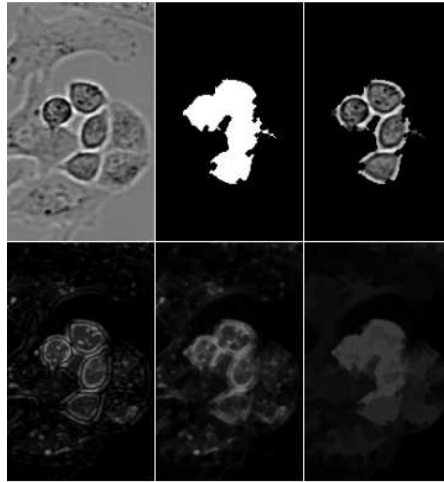
The parameters used in the proposed algorithm belongs to two groups. First contains all morphological operators described by Eqs.12,13,14. Parameters of the first group depends on the resolution of the image and the average size of cells in pixels. The second group to which the double thresholding (Eq.15) belongs, requires graylevel thresholds which depend on the luminance of cells. The latter, in turn, depends on the lighting conditions under which the initial image was taken. As far as the first group is concerned, structuring elements shapes and sizes have to be defined. They depend on the size of the average cell. All the structuring elements are described in Table 1. All these structuring elements was chosen for input image of resolution 1024x1024.

Parameter from the second group - thresholds in Eq.15 depends on the intensity of the image. They was chosen manually and were constant for all the images within single series.

Results of the proposed method of converged cells fraction localization are presented in Fig. 3.

Table 1. Structuring elements

SE	Eq.	type
B1	12	Octagon of radius 6
B2	12	Octagon of radius 3
B3	13	4-connected elementary
B4	14	Octagon of radius 3
B5	14	Octagon of radius 6
BR	12,13,14	8-connected elementary

**Fig. 3.** Example of cell extraction

6 Conclusions

The method of extraction cell fraction which *in potentia* are mitotic cells has been proposed in the paper.

The method is based on the morphological image processing. It consist of three steps: background suppression, image simplification and binarization. In each of these steps advanced morphological tools was applied mostly based on the morphological reconstruction. The application of these tools allows dealing with complex images of cells. These images are characterized by several features which causes serious problems for automatic processing. First, there are cells of various kinds determined by their shape, intensity and texture in each image. Secondly, the image acquisition conditions varies from one series of images to the others. Consequently, images, although presenting biological material of homogeneous type, are themselves not homogeneous. This fact caused a problem for automatic extraction of desired cell type. Proposed algorithm deals well with all above problems, which was illustrated in the paper by images showing

processing results. The method extracting regions occupied by cells which appear like cells before division on most of processed images. It does not answer if cells are mitotic or not but feather cells behaviors monitoring using next images in sequence allows to determinate it.

Future research will focus on the ability of the method to detect other types of cells, as well as the detection of single cells instead of regions occupied by cells in clusters.

Acknowledgement. We are grateful for confocal microscope images of neural stem cells in culture from the NeuroRepair Department, Polish Academy of Sciences Medical Research Centre, Warsaw, Poland.

References

1. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., Watson, J.D.: *Molecular Biology of the Cell*, 3rd edn. Garland Publishing Inc., New York, London (1994)
2. Boucher, A., Doisy, A., Ronot, X., Garbay, C.: Cell Migration Analysis After in Vitro Wounding Injury with a Multi-Agent Approach. *Artificial Intelligence Review* 12, 137–162 (1998)
3. Boier Marti, I.M., Martineus, D.C., et al.: Identification of spherical virus particles in digitized images of entire electron micrographs. *Journal of Structural Biology* 120, 146–157 (2005)
4. Buzanska, L., Jurga, M., Stachowiak, E.K., Stachowiak, M.K., Domanska-Janik, K.: *Stem Cell and Development*, vol. 15, pp. 391–406 (2006)
5. Francis, K., Ramakishna, R., Holloway, W., Palsson, B.O.: Two New Pseudopod Morphologies Displayed by the Human Hematopoietic KG1a Progenitor Cell Line and by Primary Human CD34+ Cells. *Blood* 92(10), 3616–3623 (1998)
6. Frank, J., Radermacher, M., et al.: Spider and web: Processing and visualization of images in 3d electron microscopy and related fields. *Journal of Structural Biology* 116, 190–199 (1996)
7. Iwanowski, M.: Binary Shape Characterization Using Morphological Boundary Class Discrimination Functions. In: Kurzynski, M., Puchala, E., Wozniak, M., Zolnierek, A. (eds.) *Computer Recognition Systems*, pp. 303–312. Springer, Heidelberg (2007)
8. Kivioja, T., Ravantti, J., et al.: Local average intensity-based method for identifying spherical particles in electron micrographs. *J. of Structural Biology* 131, 126–134 (2000)
9. Korzyńska, A., Strojny, W., Hoppe, A., Wertheim, D., Hoser, P.: Segmentation of microscope images of living cells. *Pattern Anal. Applic.* 10, 301–319 (2007)
10. Korzyńska, A.: Automatic Counting of Neural Stem Cells Growing in Cultures. In: Kurzynski, M., Puchala, E., Wozniak, M., Zolnierek, A. (eds.) *Computer Recognition Systems*, pp. 604–612. Springer, Heidelberg (2007)

11. Korzyńska, A.: Neutrophils' movement in vitro. *Annals of The New York Academy of Sciences* 972, 139–143 (2002)
12. Ludtke, J., Baldwin, P., Chiu, W.: EMAN: Semiautomated software for high-resolution signal-particle reconstruction. *J. of Structural Biology* 128, 82–97 (1999)
13. Miroslaw, L., Chorazyczewski, A., Frank, B., Kittler, R.: Correlation-based Method for Automatic Mitotic Cell Detection in Phase Contrast Microscopy. In: Kurzynski, M., Puchala, E., Wozniak, M., Zolnierek, A. (eds.) *Computer Recognition Systems*, pp. 627–634. Springer, Heidelberg (2005)
14. Nicholson, W.V., Glaeser, R.M.: Review: Automatic particle detection in electron microscopy. *Journal of Structural Biology* 133, 90–101 (2001)
15. Nieniewski, M.: *Morfologia matematyczna w przetwarzaniu obrazów*, PLJ Warszawa (1998)
16. Russ, J.C.: *Image Processing Handbook*, 4th edn. CRC Press, Tokyo (2002)
17. Serra, J., Vincent, L.: An overview of morphological filtering, *Circuit systems Signal Processing*, vol. 11(1) (1992)
18. Serra, J.: *Image analysis and mathematical morphology*, vol. 1. Academic Press, London (1983)
19. Serra, J.: *Image analysis and mathematical morphology*, vol. 2. Academic Press, London (1988)
20. Smereka, M.: Detection of ellipsoidal shapes using contour grouping. In: Kurzynski, M., Puchala, E., Wozniak, M., Zolnierek, A. (eds.) *Computer Recognition Systems*, pp. 443–450. Springer, Heidelberg (2005)
21. Soille, P.: *Morphological image analysis*. Springer, Heidelberg (2002)
22. Solomon, E.P., Berg, L.R., Martin, D.W.: *Biologia MULTICO* oficyna Wydawnicza, Warszawa (2007)
23. Vincent, L.: Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithms. *IEEE Trans. on Image Processing* 2(2) (1993)
24. Zama, N., Katow, H.: A method of quantitative analysis of cell migration using a computerized time-lapse videomicroscopy. *Zool. Sci.* 5, 53–60 (1988)

Protein Molecular Viewer for Visualizing Structures Stored in the PDBML Format

Dariusz Mrozek, Andrzej Mastej, and Bożena Małyśiak

Institute of Informatics, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland

Dariusz.Mrozek@polsl.pl, Bozena.Malyśiak@polsl.pl, andrewus@interia.pl

Summary. Visualization of protein molecular structures is a very important part of the analysis of protein function and activity. Visualization tools allow to represent graphically the complex construction of proteins and give us an idea of the biological molecules built with hundreds or thousands of atoms linked to each other by covalent bonds. In the chapter we present the most interesting features of our Protein Molecular Viewer (PMV). The PMV is a molecular visualization tool which is used to show protein structures loaded from the well-known Protein Data Bank. With the possibility of loading and presenting protein structures from the PDBML data format the PMV becomes one of a few tools in the world having this unique function.

1 Introduction

Applications that visualize the spatial structures of proteins and other biological compounds belong to the wide group of tools of molecular analysis used in biochemistry, proteomics and system biology. Functioning of living organisms in biological aspect is tightly related with the existence and activity of proteins. Proteins are important molecules that play a key role in all biochemical reactions in organisms' cells. They are involved in many processes, e.g.: reaction catalysis, energy storage, signal transmission, maintaining of cell mechanical structure, immune response, stimuli response, cellular respiration, transport of small biomolecules, regulation of cell growth and division [1, 2].

Analyzing their general construction proteins are macromolecules with the molecular mass above 10 kDa ($1 \text{ Da} = 1.66 \times 10^{-24} \text{g}$) built with amino acids (>100 amino acids, aa). Amino acids are linked in linear chains by peptide bonds [3]. In the construction of proteins we can distinguish four description (or representation) levels: primary structure, secondary structure, tertiary structure and quaternary structure. The last three levels define the protein conformation or protein spatial structure [3, 4]. The biochemical analysis is usually carried on one of the description levels.

The analysis of protein spatial structure is very important from the viewpoint of protein function, protein activity and reactions the protein is involved in. This type of analysis supported by the observations of protein structure, include not only a sequence, but also geometrical features of studied molecule. There is no

doubt, the structures of even small molecules are very complex – proteins are built up of hundreds of amino acids, and then thousands of atoms. Visualization tools, which allow to display and to study spatial structures in the finest details, are useful to explore such complex structures [5]. The common purpose of the tools is a presentation of the general atomic structure of proteins, general spatial shape, and presentation of the simplified construction by the extraction and revealing of secondary structures. As a result, a user can study the shape of a protein or its specific regions and compare it to other proteins, and thus evaluate the similarities and differences. Since the early eighties scientists have made a use of growing knowledge about various protein structures and functions to rebuild existing proteins and to design completely new molecules. For this reason, protein structure viewers are often applied in drug engineering - they help in the design of effective drugs. The achievements of the modern pharmacy are impressive. However, it is impossible to construct a new drug until the structure of the pathogenic, malfunction protein is found out. Visualization tools are then supportive in the design of inhibitors - substances that decreases excessive activity of some proteins, especially enzymes [6]. One of a vibrant branch of the modern biochemistry, molecular biology and biotechnology became the prediction of protein structures, since it gives many possibilities to the medicine and industry. This process cannot be performed without the insight into protein internal arrangement. Finally, visualization tools are indispensable in the molecular pathology, where scientists investigate the influence of small mutations in protein structures on the protein activity [4].

The visualization of protein structures is performed on the basis of structural data, which have a form of Cartesian coordinates (x, y, z) . These coordinates can be retrieved from appropriate databases. Therefore, the visualization of proteins is possible on personal computers. The most popular databases are Protein Data Bank (PDB) [7] and NCBI Molecular Modeling DataBase (MMDB) [8]. They contain data obtained as a result of X-ray crystallography or NMR spectroscopy. These repositories make the data available for an exchange in appropriate formats, like: PDB [9], mmCIF [10] and PDBML [11] for the PDB repository, and ASN.1 [12] for the MMDB repository. The PDBML [11] is the newest format, which benefits from the strength of the XML technology and it will certainly become the main exchange format of structural data for the Protein Data Bank (PDB).

In the chapter, we present the Protein Molecular Viewer (PMV) - a tool to visualize protein structures stored in the PDBML data sets. The PMV has several unique features that distinguish it from other viewers. These applications are briefly described in section 2. The most important functions of the Protein Molecular Viewer are demonstrated in section 3.

2 Related Works

Computer scientists have developed protein visualization tools for many years. These tools are able to show protein structures stored in many different formats

designed during all these years. However, in the section we give a short description of the tools in the context of the exchange formats related to the PDB and MMDB repository.

RasMol [13] is one of the most popular programs for the visualization of molecular structures and chemical compounds. It can read protein structures from the PDB and mmCIF formats. The RasMol enables many representations of protein structure and has many functions. Users can zoom in/out and rotate obtained structures and mark particular groups of atoms with a chosen color. In the RasMol it is also possible to export obtained pictures to popular graphical formats. The program was implemented in the C language and its graphical performance is very good.

The Jmol [14] is a tool realized in the Java language, which visualizes structures stored in the PDB format (and some other, not related to the PDB). The functionality of this program is comparable to the RasMol. A graphic engine used for 3D visualization was written in the Java language. This engine was optimized for molecule display. The performance of this program is very good although it was written in the Java language. The main disadvantage of the Jmol is that there is no possibility to read structures stored in the PDBML format.

The jV [15] is another known program for 3D visualization of proteins. The program was written in the Java language and uses the JOGL (Java bindings for OpenGL) library, which supports hardware acceleration. The jV is one of a few programs that have a possibility to read structures from the newest PDBML format. The program provides a very efficient visualization and its functionality is comparable to RasMol.

The Cn3D [16] is a very efficient viewer of protein 3D structures, implemented in the C++ language. The program uses the OpenGL technology. The quality of graphical representation is very high at very fast rendering. However, the Cn3D can visualize proteins from the MMDB repository, which are stored in the ASN.1 format [12].

The Bioclipse [17] is an interesting tool for 3D visualization of biological molecules dedicated for chemists and bioinformatics specialists. The program was realized in the Java language based on the Eclipse Rich Client Platform (RCP). The Bioclipse inherits a basic functionality and visual interfaces from the Eclipse environment, which functionality can be extended by additional plug-ins. For the 3D visualization of molecules a user should install the Jmol plug-in. However, the functionality of the Jmol plug-in is very limited in comparison to the Jmol application presented above. The Bioclipse can visualize different kind of chemical compounds and also their chemical formula.

3 The Protein Molecular Viewer

The Protein Molecular Viewer (PMV) is a program, which visualize molecular structures of proteins and other compounds. It was implemented in the Java language and can work as an application or the Java applet on any website. The PMV can read molecular data in the PDBML format, usually retrieved from the

PDB repository. Older formats, like PDB or mmCIF, are not supported, since the PMV concentrates on web technologies and popular XML format. The PMV uses the Java3D library to display spatial structures of biological molecules. The library has to be installed before the PMV is executed. The PMV program works in the specific architecture, which makes up its strength and will be presented in the next section.

3.1 PMV Work Architecture

The PMV program works with data sets, which can be stored locally or distributed over the Internet (Fig. 1). Therefore, the PMV users can open and visualize molecular structures saved in the PDBML format on the local hard drive or download structures from the web. In the second case, molecular structures can be downloaded directly from the XML repository of the Protein Data Bank in the United States or from any place identified by HTTP or FTP address (e.g. <http://www.pdb.org/pdb/files/2abx.xml>). As a result, the PMV can visualize structures that are the elements of various network resources. The only condition is that molecular structures must be kept in the form of PDBML files. Users usually specify the PDB ID number (PDB identifier) of the structure, which they want to view.

3.2 Display Modes

During the visualization of protein structures or other chemical compounds users have their own preferences regarding structure representation. It usually depends on the visualization purpose, e.g. is it a profound analysis of the structure concerning the smallest details? or, is it just a preview of the general structure of the molecule? For this reason, the PMV program provides three basic display modes:

- atomic - it is the most detailed display mode, which shows all atoms in a structure and bonds between atoms (Fig. 2a). The atomic mode focuses on

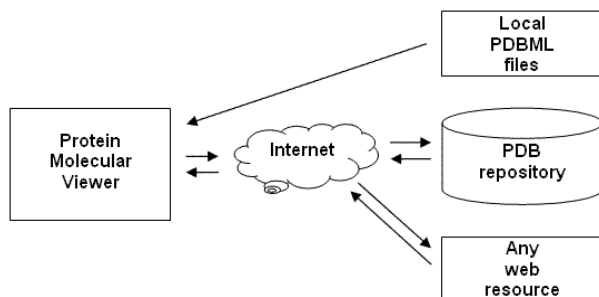


Fig. 1. The architecture of the data exchange during the structure visualization with the use of the Protein Molecular Viewer

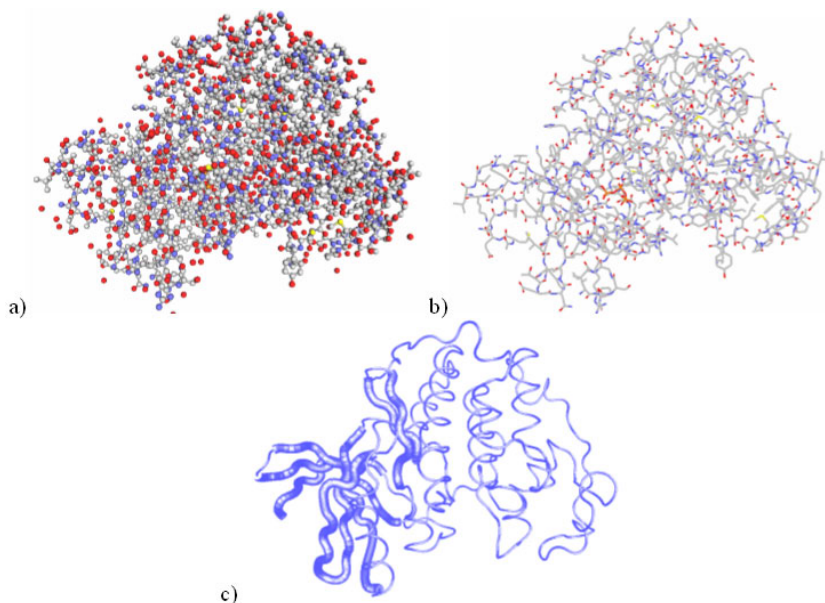


Fig. 2. The structure of the CDK2 kinase (PDB ID: 1B38) displayed by the PMV: a) atomic mode, b) sticks mode, c) ribbon mode

atoms, which can be supportive when users evaluate a distance between chosen atoms or observe small conformational switching as an effect of particular cellular reaction. However, a consequence of the big number of details in the atomic mode is that usually only the elements of the first scene are visible.

- sticks (frame) - focuses on the skeleton of the displayed molecule, which is formed by covalent interatomic bonds (Fig. 2b); The atoms themselves are marked with particular colors at the ends of the bonds. The sticks mode is less detailed than atomic mode. The mode is especially supportive for observations of the conformation changes as a consequence of various factors.
- ribbon - reveals the secondary structure elements in the construction of the molecule (Fig. 2c); It is the least detailed display mode, recommended to the analysis of the general structure of the biological molecules, e.g. in protein function identification or visual evaluation of the protein similarity.

3.3 Marking Selected Structural Regions

One of the interesting features of the PMV is a possibility to mark selected structural regions of the molecule through the change of its color or selection of the particular group of atoms. The PMV provides two predefined approaches to painting displayed structures: atom painting (default) and chain painting.

Atom painting is only available in the atomic and sticks display modes. It distinguishes atoms depending on the chemical element. In the PMV we assumed

Table 1. Colors of chemical elements in the atom painting

Chemical element	Symbol	Color	Chemical element	Symbol	Color
Oxygen	O	red	Sulfur	S	yellow
Hydrogen	H	white	Phosphor	P	orange
Nitrogen	N	blue	Others	-	green
Carbon	C	light gray			

coloring of atoms according to the rules presented in table 1. The atom painting, as a default setting, was visible in Fig. 2a and Fig. 2b.

Chain painting is available in all display modes. It distinguishes individual chains in the quaternary structure of a protein. Chain painting in various display modes is presented in Fig. 3.

The third option of the structure painting is a marking only selected group of atoms in the displayed structure. Selected atoms are then marked with a chosen color (Fig. 4). It is the most advanced form of the structure painting. This painting option is not predefined and is very useful during the analysis of the particular regions of proteins. E.g. if a user wants to view, how the active site of an enzyme is constructed, he/she can select and mark with a chosen color only these atoms, which belong to the active site. Afterwards, it is much easier to identify particular atoms in the complex structure displayed at the screen, zoom the image in and watch selected atoms carefully or evaluate distances between

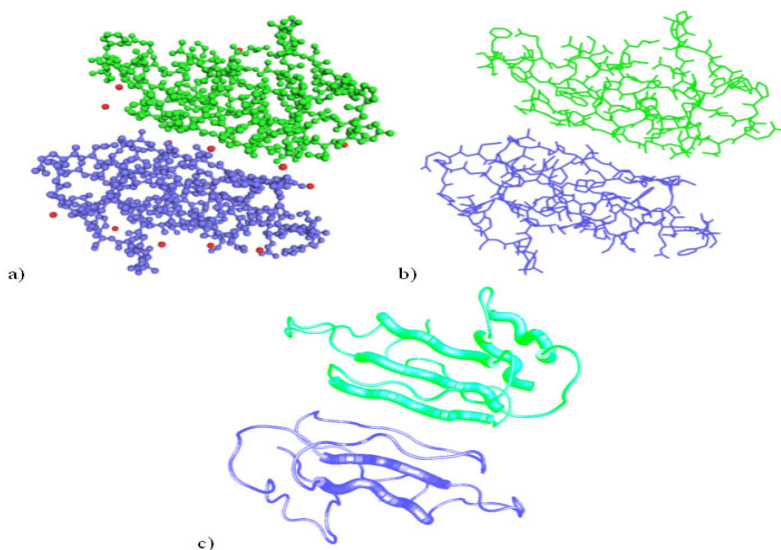


Fig. 3. Chain painting in the PMV for the Alpha-Bungarotoxin molecule (PDB ID: 2ABX): a) atomic mode, b) sticks mode, c) ribbon mode

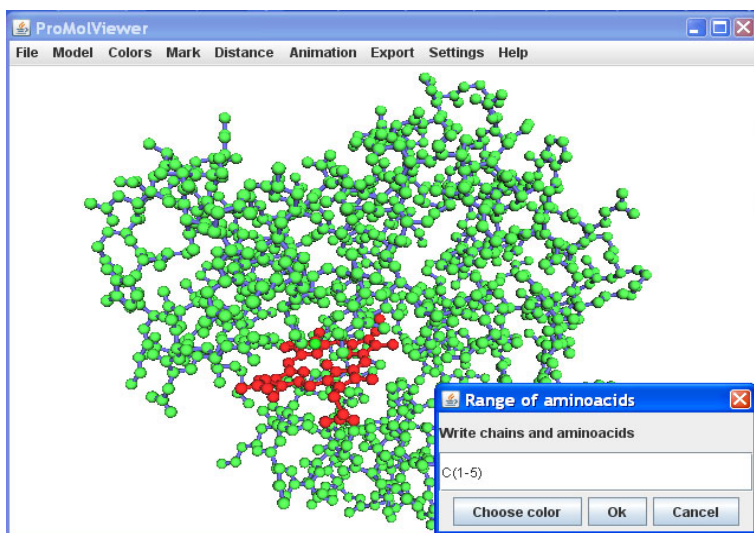


Fig. 4. Structure painting by marking a group of atoms - part of the structure of the myoglobin (PDB ID: 1MBN) responsible for oxygen binding

each other. Marking a group of atoms the user specifies chains and amino acids (or amino acid collections) the requested atoms belong to. E.g. a notation A(1-15) means the user wants to mark all atoms from amino acids 1 to 15 in the chain A of the presented molecule. He/she can also write more complex expressions, e.g. A(1-15);B(10);C(20-35), which allow to mark several disjoint groups of atoms.

3.4 Additional Features of the PMV

The PMV has many other interesting features. However, only some of them will be described in the section. One of the useful features of the PMV is that it allows to evaluate distances between pointed pairs of atoms. A user can do this in the atomic display mode. The information about the distance is shown in additional window and straight line is drawn between selected atoms (Fig. 5). The distance is given in Angstroms ($1\text{A} = 1 \times 10^{-10}\text{m}$). This kind of measurements is supportive for the analysis of selected parts of molecular structures.

During the work with the PMV, a user can manipulate the 3D scene by using a mouse: rotate a structure (by moving the pointer while keeping the left mouse button pressed), displace a structure (by moving the pointer while keeping the right mouse button pressed), zoom in/out (by moving the pointer down/up while keeping the middle mouse button or ALT key pressed).

A user is allowed to export displayed structure in the current view to one of the popular graphic formats. In the PMV the following formats are supported: JPG, PNG, BMP, and GIF. If the quality of the view is insufficient, a user can change the PMV display resolution and the background color.

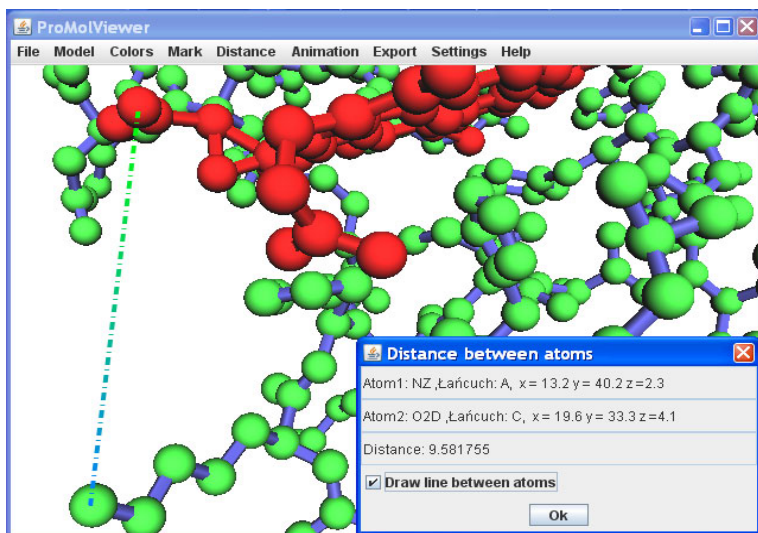


Fig. 5. The PMV: Marked and enlarged part of the myoglobin structure (PDB ID: 1MBN) responsible for oxygen binding with the line evaluating the distance between two selected atoms

3.5 Software and Hardware Requirements

It is required to install the Java Runtime Environment (JRE) and Java 3D library before the first use of the PMV. The PMV is optimized to work with the JRE ver. 6 update 3 and Java 3D ver. 1.5.0. Both, the JRE and Java 3D can be downloaded from the Internet. The Java 3D library should be installed manually, if the PMV is launched as an application, or will be downloaded and installed automatically, if the PMV is launched as an applet. The PMV applet has a digital signature - a user will be asked to accept the signature before the use of applet on the www web site. The PMV applet was tested and works with the following web browsers: MS Internet Explorer ver. 6 and 7, and Mozilla Firefox.

The required item of the hardware is a graphics card with the 3D graphics acceleration and the support for the OpenGL technology ver. 1.2.

3.6 The PMV Software Availability

The Protein Molecular Viewer is free for all, who need a tool to visualize structures from the Protein Data Bank stored in PDBML format files. The PMV supports two languages: English and Polish. A user can easily switch between languages of the PMV GUI. The PMV and required libraries can be downloaded from the Protein Molecular Viewer Home Page <http://zti.polsl.pl/dmrozek/pmView.htm>

4 Concluding Remarks

The Protein Molecular Viewer is still in the development phase. However, the first version of the tool has already had unique features in comparison to other molecular viewers. One of the main advantages is a possibility to load and show data stored in the newest PDBML format. At the moment, only a few tools in the world support this format. There were other important functions implemented in the PMV, like: structure painting, various display modes, an interatomic distance evaluation, export of the image to popular graphics formats, listing of additional information about the displayed structure, rotations, moving, scaling, animation of the structure, and others. These functions are very practical and valuable for the structural analysis of proteins. We use the PMV in our research to verify the results of the protein similarity search processes performed by our EAST algorithm [18, 19]. However, the purpose and functionality of the program is much wider. A significant feature of the PMV is an ability to download and display protein structures directly from the Protein Data Bank repository or from any web resource using the HTTP or FTP protocols. This is a very rare ability among visualization tools. The PMV was implemented in the Java language, and consequently, it is independent on the hardware and system platform. Moreover, it can work as an applet, which can be a part of any web site.

The PMV has also some weak points. The main disadvantage concerns efficiency decrease during the visualization of big molecular structures. This leads to a big consumption of the memory. The sticks display mode is the least, and the atomic display mode is then the most memory consuming. The main reason of the decreasing efficiency is a rising complexity of the 3D scene built by Java 3D library. A solution of the problem is planned to be an element of the future works on extending the PMV functionality.

References

1. Fersht, A.: *Enzyme Structure and Mechanism*, 2nd edn. W.H. Freeman & Co., New York (1985)
2. Dickerson, R.E., Geis, I.: *The Structure and Action of Proteins*. 2nd edn. Benjamin/Cummings, Redwood City, Calif. Concise (1981)
3. Hames, B.D., Hooper, N.M., Houghton, J.D.: *The Instant Notes in Biochemistry*. BIOS Scientific Publishers Ltd., Oxford (1997)
4. Creighton, T.E.: *Proteins: Structures and molecular properties*, 2nd edn. Freeman, San Francisco (1993)
5. Alberts, B., Johnson, A., Lewis, J., Raff, M., et al.: *Molecular biology of the cell*, 4th edn., Garland Science, NY (2002)
6. Burkert, U., Allinger, N.L.: *Molecular Mechanics*. American Chemical Society, Washington D.C (1980)
7. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., et al.: The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242 (2000)
8. Marchler-Bauer, A., Address, K.J., Chappay, C., Geer, L., et al.: MMDB: Entrez's 3D structure database. *Nucleic Acids Res.* 27(1), 240–243 (1999)

9. Westbrook, J.D., Fitzgerald, P.M.D.: The PDB Format, mmCIF Formats, and Other Data Formats. In: Structural Bioinformatics, vol. 44, John Wiley & Sons, Inc., Chichester (2003)
10. Bourne, P.E., Berman, H.M., Watenpaugh, K., et al.: The macromolecular Crystallographic Information File (mmCIF). *Methods Enzymol.* 277, 571–590 (1997)
11. Westbrook, J., Ito, N., Nakamura, H., Henrick, K., Berman, H.M.: PDBML: the representation of archival macromolecular structure data in XML. *Bioinformatics* 21(7), 988–992 (2005)
12. Ohkawa, H., Ostell, J., Bryant, S.: MMDB: an ASN.1 specification for macromolecular structure. In: Proc. Int. Conf. Intell. Syst. Mol. Biol., vol. 3, pp. 259–267 (1995)
13. Sayle, R.: RasMol, Molecular Graphics Visualization Tool. Biomolecular Structures Group, Glaxo Wellcome Research & Development, Stevenage, Hertfordshire (1998)
14. Steinbeck, C., Han, Y., Kuhn, S., Horlacher, O., et al.: The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemo- and Bioinformatics. *Journal of Chemical Information and Computer Sciences* 43(2), 493–500 (2003)
15. Kinoshita, K., Nakamura, H.: jV - Documentation and Users Guide (2008), <http://www.pdbj.org/jV/Help.html>
16. Hogue, C.W.: Cn3D: a new generation of three-dimensional molecular structure viewer. *Trends Biochem. Sci.* 8, 314–316 (1997)
17. Spjuth, O., Helmus, T., Willighagen, E.L., Kuhn, S., et al.: Bioclipse: an open source workbench for chemo- and bioinformatics. *BMC Bioinformatics* 8, 59 (2007)
18. Mrozek, D., Małysiak, B., Kozielski, S.: EAST: Energy Alignment Search Tool. In: Wang, L., Jiao, L., Shi, G., Li, X., Liu, J. (eds.) FSKD 2006. LNCS (LNAI), vol. 4223, pp. 696–705. Springer, Heidelberg (2006)
19. Mrozek, D., Małysiak, B.: Searching for Strong Structural Protein Similarities with EAST. *Journal of Computer Assisted Mechanics and Engineering Sciences* 14, 681–693 (2007)

Fuzzy Support Vector Machine for Genes Expression Data Analysis

Joanna Musioł¹, Agnieszka Więclawek², and Urszula Mazurek²

¹ Silesian University of Technology, Department of Biomedical Engineering, Gliwice, Poland

jmusiol@polsl.pl

² Silesian Medical University, Department of Molecular Biology and Medical Genetics, Sosnowiec, Poland

Summary. The current study presents two approaches to the fuzzy support vector machine. The first approach implements the fuzzy support vector machine for solving a two class problem. The second approach employs the fuzzy support vector machine for a multi-class problem. In both cases fuzzy classifiers have been used for genes expression data analysis. The first method has been tested on clinical data acquired at the Silesian Medical University. Then the dataset from Kent Ridge Biomedical Data Set Repository has been used to simulate the performance of the second tool.

1 Introduction

The support vector machines (SVM) is a classification technique based on the statistical learning theory, first introduced by Vapnik in 1995. A good classification performance of this tools is commonly known for many years.

Many approaches to the biomedical data analysis have been developed, yet no general method has been found. SVM combined with fuzzy logic (FSVM) is an effective tool which opens up new data classification possibilities.

This paper is organized as follows. In Sec. 2 a definition of the classical approach to support vector machines is given. Section 3 discusses a modification of the classical FSVM-theory for two classes. The FSVM for a multi-class problem is described in Sec. 4. Section 5 discusses two datasets and results of the experiment. The last section (Sec. 6) concludes the paper.

2 Support Vector Machine

In a classical approach the SVM is a learning machine able to solve a two-group classification problem. This method permits for an optimal hyperplane data separation. Optimal hyperplane is a unique hyperplane which separates the training data with a maximal margin [2]. For the linearly separable labeled training set

$$(x_1, y_1), \dots, (x_l, \dots, y_l), y_i \in \{-1, 1\} \quad (1)$$

a vector \mathbf{w} (weight vector) and a scalar \mathbf{b} (bias or threshold) fulfill the inequalities

$$\mathbf{w} \cdot \mathbf{x}_i + b \geq 1 \quad \text{if } y_i = 1 \tag{2}$$

$$\mathbf{w} \cdot \mathbf{x}_i + b \leq -1 \quad \text{if } y_i = -1 \tag{3}$$

The separating hyperplane is given by

$$\mathbf{w}^T \mathbf{x} + b = 0 \tag{4}$$

with the decision rule given by

$$f_{\mathbf{w},b}(\mathbf{x}) = \text{sgn}(\mathbf{w}^T \mathbf{x} + b). \tag{5}$$

Vectors \mathbf{x}_i for which $y_i(\mathbf{w}\mathbf{x}_i + b) = 1$ are referred to as **support vectors**. The optimal hyperplane is obtained by solving the following quadrating programming problem

$$\text{Minimize}_{\mathbf{w},b} \Phi(\mathbf{x}) = \frac{1}{2} \|\mathbf{w}\|^2 \tag{6}$$

$$\text{s.t.} \quad y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad i = 1, \dots, l. \tag{7}$$

This problem is precisely described in [2].

The biomedical data is often not linearly separable. In order to use the SVM approach for solving the nonlinear problem, a training-data has to be first mapped to a higher dimensional **features space**. The separating hyperplane is constructed in this space. A function that transforms the n-dimensional input vector into a N-dimensional feature vector $\phi : R^n \mapsto R^N$ is called a kernel function. The classification of an unknown vector \mathbf{x} takes a form

$$f_{\mathbf{w},b}(x) = \text{sgn}(\mathbf{w}^T \phi(\mathbf{x}) + b) \tag{8}$$

In this papers three types of kernels are tested:

Linear

$$K(\mathbf{u}, \mathbf{v}) = \mathbf{u}^T \cdot \mathbf{v} \tag{9}$$

Polynomial

$$K(\mathbf{u}, \mathbf{v}) = (\langle \mathbf{u}, \mathbf{v} \rangle + 1)^k \quad \text{k-degree of a polynomial} \tag{10}$$

Gaussian Radial Basis Function

$$K(\mathbf{u}, \mathbf{v}) = \exp\left(-\frac{\|\mathbf{u} - \mathbf{v}\|}{2\sigma^2}\right) \tag{11}$$

3 Fuzzy Support Vector Machine for Two Classes

Randomly chosen SVMs can by a reason of misclassifications occurrence. One solution is to classify the same data of different SVMs and use fuzzy knowledge to combine the classifications results [3].

To make the calculation easier the same data is classified only with three SVM classifiers. The classifiers could be different in types and parameters of kernel functions. Three classifiers are trained and validation data examples are then plugged into the decision hyperplane models. (Validation data examples means the part of training data, which is not used for training the models.) Accuracies of these three different classifiers are then obtained. Next, the set of testing data is subjected to the classification of the three previously used models. As a classification result, for each testing point the distances \mathbf{d} from these optimal hyperplanes are calculated as:

$$d(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b. \tag{12}$$

Accuracies and distances are then used in the next phase [3], where fuzzy system is constructed. The inputs of the fuzzy system are accuracies $a_i, i = 1, \dots, 3$ and distances $d_i, i = 1, \dots, 3$. For the sake of the need of simplifying the calculation, triangles membership functions are employed. These functions are shown in Fig.1 [3]. Therefore, fuzzy sets for all inputs consists of 216 rules, which combine accuracies and distances as follows:

In a_1 is A_1 and a_2 is A_2 and a_3 is A_3 and d_1 is D_1 and d_2 is D_2 and d_3 is D_3 then

$$z = \sum_{i=1}^3 \mu_{A_i}(a_i) \mu_{D_i}(d_i) \lambda_i d_i a_i v_i \tag{13}$$

where $A_1, A_2, A_3 \in \{low, middle, high\}$ and $D_1, D_2, D_3 \in \{negative, positive\}$, $\mu_{A_i}(a_i)$ and $\mu_{D_i}(d_i)$ are membership functions. λ denotes an adjustment factor of the fuzzy models [3], whose value depends on the accuracy: if the accuracy is greater than or equal to 80%, then $\lambda = 1$; if the accuracy ranges between 50% and 80% then $\lambda = \frac{1}{2}$ and is equal to $\frac{1}{4}$ when the accuracy is lower than 50%. v_i describes observation applied in kernel function, discussed in Sec. 5.

To determine the firing strength of one particular rule T-norm, the min operator is used as follows [3]:

$$\eta_i = \mu_{A_1}(a_1) \wedge \mu_{A_2}(a_2) \wedge \mu_{A_3}(a_3) \wedge \mu_{D_1}(d_1) \wedge \mu_{D_2}(d_2) \wedge \mu_{D_3}(d_3) \tag{14}$$

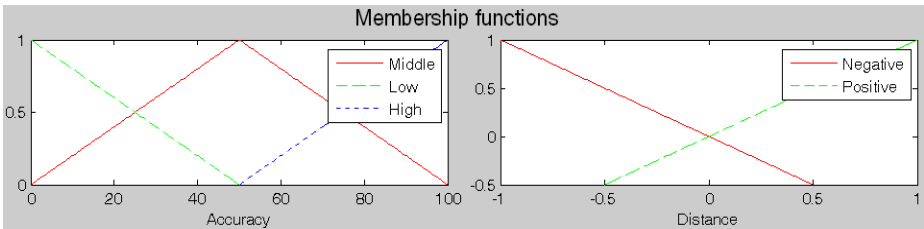


Fig. 1. Membership functions [3]

The final result is given as:

$$Z' = \sum_{i=1}^K \eta_i z_i / \sum_{i=1}^K \eta_i, \tag{15}$$

where K is the number of rules. If Z' is greater than or equal to zero, the testing data point is considered in the positive class. Otherwise, it belongs to the negative class [3].

4 FSVM for Multi-class Problem

The classical SVM handles a two-class problem. In order to use the conventional method for solving a multi-class problem an n -class problem has to be converted into n two-class problems and for the i th two-class problem, class i is separated from the remaining classes [4]. This method is called One-against-One (1A1) [5]. Another approach is based on [6] converting the n -class problem into $n(n-1)/2$ two-class problems. This method is called One-against-All (1AA) [5] or pairwise classification. However, unclassified regions remain when any of these methods is used. To solve this problem a fuzzy support vector machine for multi-class problems is developed.

4.1 One-against-One

In conventional pairwise classification, the decision function for class i against class j , with the maximum margin is formulated as follows:

$$D_{ij}(\mathbf{x}) = \mathbf{w}_{ij}^T \phi(\mathbf{x}) + b_{ij} \tag{16}$$

where $D_{ij}(\mathbf{x}) = -D_{ji}(\mathbf{x})$.

For each input vector \mathbf{x} the decision function is calculated as [6]:

$$\arg \max_{i=1, \dots, n} D_i(\mathbf{x}) \tag{17}$$

where

$$D_i(\mathbf{x}) = \sum_{i \neq j, j=1}^n \text{sgn} D_{ij}(\mathbf{x}). \tag{18}$$

A problem appears if equation (18) satisfies more than one class. In this case \mathbf{x} is unclassifiable. For solving this problem a one-dimensional membership function $\mu_{ij}(\mathbf{x})$ on the directions orthogonal to $D_{ij}(\mathbf{x}) = 0$ is defined as:

$$\mu_{ij}(\mathbf{x}) = \begin{cases} 1 & \text{for } D_{ij}(\mathbf{x}) \geq 1 \\ D_{ij}(\mathbf{x}) & \text{otherwise} \end{cases} . \tag{19}$$

Since $\mu_i(\mathbf{x}) = 1$ holds for one class only

$$\mu_i(\mathbf{x}) = \min_{j=1, \dots, n} D_{ij}(\mathbf{x}) \quad \text{for } i \neq j. \tag{20}$$

A new vector \mathbf{x} is classified into the class

$$\arg \max_{i=1, \dots, n} \mu_i(\mathbf{x}). \tag{21}$$

4.2 One-against-All

Similar to the formulation in Sec. 4.1, a solution of a N two-class problem is defined. The decision function that classifies a class i and separates i from the remaining classes is given by [4]:

$$D_i(\mathbf{x}) = w_i^T \phi(\mathbf{x}) + b_i, \tag{22}$$

if

$$\text{sgn}(D_i(\mathbf{x})) = 1 \tag{23}$$

where \mathbf{x} belongs to the i th class. If equation (23) is satisfied for various cases, \mathbf{x} is unclassifiable. Then, a one dimensional membership function for each class $\mu_{ij}(\mathbf{x})$ is given as:

$$\mu_{ii}(\mathbf{x}) = \begin{cases} 1 & \text{for } D_i(\mathbf{x}) > 1 \\ D_i(\mathbf{x}) & \text{otherwise} \end{cases} \tag{24}$$

for $i \neq j$

$$\mu_{ij}(\mathbf{x}) = \begin{cases} 1 & \text{for } D_j(\mathbf{x}) < -1 \\ -D_j(\mathbf{x}) & \text{otherwise} \end{cases} \tag{25}$$

For the class i the membership function is defined as:

$$\mu_i(\mathbf{x}) = \min_{j=1, \dots, n} \mu_{ij}(\mathbf{x}), \tag{26}$$

therefore, the \mathbf{x} is classified into the class

$$\arg \max_{i=1, \dots, n} \mu_i(\mathbf{x}). \tag{27}$$

5 Experimental Results

In this section experimental results of applying the fuzzy support vector machine and the conventional support vector machine to a gene expression analysis are presented.

5.1 Two-Classes

The first experiment has been carried out on the microarray of the genes connected with dopamine metabolism in normal placentas and placentas suffering with hypertension (PIH) and diabetes (GDM) [7]. A set of probes used in the experiment has been collected at the Department and Clinic of Perinatology and Gynecology Medical University of Silesia in Zabrze. It consists of samples from patients with PIH (7 cases), with GDM (5) and from control group (3). The data has been classified on the basis of distinguished transcripts previously selected by Bland-Altman method. Two transcripts have been distinguished between PIH and control samples, and two other between GDM and control samples.

The default parameters of SVMs have been used to train the datasets as shown Table 1. For PIH and control samples 10 tests has been carried out. In

Table 1. Parameters of used SVMs

Nr	Parameters of SVMs					
	SVM1		SVM2		SVM3	
	kernel	parameters	kernel	parameters	kernel	parameters
1	RBF	$\sigma = 0.1$	RBF	$\sigma = 1$	RBF	$\sigma = 10$
2	RBF	$\sigma = 1$	RBF	$\sigma = 10$	RBF	$\sigma = 100$
3	RBF	$\sigma = 1$	poly	$k = 2$	poly	$k = 3$
4	RBF	$\sigma = 10$	poly	$k = 1$	poly	$k = 5$
5	RBF	$\sigma = 10$	poly	$k = 2$	poly	$k = 3$
6	poly	$k = 1$	poly	$k = 3$	poly	$k = 5$

Table 2. Classifications result for two classes

	Nr of all cases	correct classification	improvement of result
		control group	
GDM	48	32	4
PIH	60	60	17

each of them one probe has been tested in 6 configurations of SVMs. Fig.2 lists results of the experiments. The set of GDM and control samples consists of 8 samples, therefore 8 tests in 6 configurations have been performed. In total 108 tests have been carried out. In 19.4% of cases, FSVM improves the final result and only in 5.5% of cases one of three configurations of SVM has been better than FSVM. Moreover, for FSVMs used in test PIH and control samples 100% correct classifications have been obtained. This is a very good result considering limited set of training samples taken into account. The results have been shown in Table 2.

The dataset from Kent Ridge Biomedical Data Set Repository [1] was also used for testing purposes [3]. Numerical experiment is described in [8]. For the set of all probes, in 25% cases the accuracy of FSVMs was higher than the best SVMs. In 42% cases accuracy of FSVMs was equal to the best SVMs and in

<table border="1"> <thead> <tr><th>N-35</th><th>SVM1</th><th>SVM2</th><th>SVM3</th><th>FSVM</th></tr> </thead> <tbody> <tr><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>3</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>4</td><td>1</td><td>1</td><td>0</td><td>1</td></tr> <tr><td>5</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>6</td><td>1</td><td>1</td><td>0</td><td>1</td></tr> </tbody> </table>	N-35	SVM1	SVM2	SVM3	FSVM	1	1	1	1	1	2	1	1	1	1	3	1	1	1	1	4	1	1	0	1	5	1	1	1	1	6	1	1	0	1	<table border="1"> <thead> <tr><th>N-33</th><th>SVM1</th><th>SVM2</th><th>SVM3</th><th>FSVM</th></tr> </thead> <tbody> <tr><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>3</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>4</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>5</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>6</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> </tbody> </table>	N-33	SVM1	SVM2	SVM3	FSVM	1	1	1	1	1	2	1	1	1	1	3	1	1	1	1	4	1	1	1	1	5	1	1	1	1	6	1	1	1	1	<table border="1"> <thead> <tr><th>N-24</th><th>SVM1</th><th>SVM2</th><th>SVM3</th><th>FSVM</th></tr> </thead> <tbody> <tr><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>3</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>4</td><td>1</td><td>1</td><td>0</td><td>1</td></tr> <tr><td>5</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>6</td><td>1</td><td>1</td><td>0</td><td>1</td></tr> </tbody> </table>	N-24	SVM1	SVM2	SVM3	FSVM	1	1	1	1	1	2	1	1	1	1	3	1	1	1	1	4	1	1	0	1	5	1	1	1	1	6	1	1	0	1	<table border="1"> <thead> <tr><th>N-45</th><th>SVM1</th><th>SVM2</th><th>SVM3</th><th>FSVM</th></tr> </thead> <tbody> <tr><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>3</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>4</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>5</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>6</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> </tbody> </table>	N-45	SVM1	SVM2	SVM3	FSVM	1	1	1	1	1	2	1	1	1	1	3	1	1	1	1	4	1	1	1	1	5	1	1	1	1	6	1	1	1	1
N-35	SVM1	SVM2	SVM3	FSVM																																																																																																																																											
1	1	1	1	1																																																																																																																																											
2	1	1	1	1																																																																																																																																											
3	1	1	1	1																																																																																																																																											
4	1	1	0	1																																																																																																																																											
5	1	1	1	1																																																																																																																																											
6	1	1	0	1																																																																																																																																											
N-33	SVM1	SVM2	SVM3	FSVM																																																																																																																																											
1	1	1	1	1																																																																																																																																											
2	1	1	1	1																																																																																																																																											
3	1	1	1	1																																																																																																																																											
4	1	1	1	1																																																																																																																																											
5	1	1	1	1																																																																																																																																											
6	1	1	1	1																																																																																																																																											
N-24	SVM1	SVM2	SVM3	FSVM																																																																																																																																											
1	1	1	1	1																																																																																																																																											
2	1	1	1	1																																																																																																																																											
3	1	1	1	1																																																																																																																																											
4	1	1	0	1																																																																																																																																											
5	1	1	1	1																																																																																																																																											
6	1	1	0	1																																																																																																																																											
N-45	SVM1	SVM2	SVM3	FSVM																																																																																																																																											
1	1	1	1	1																																																																																																																																											
2	1	1	1	1																																																																																																																																											
3	1	1	1	1																																																																																																																																											
4	1	1	1	1																																																																																																																																											
5	1	1	1	1																																																																																																																																											
6	1	1	1	1																																																																																																																																											
<table border="1"> <thead> <tr><th>N-42</th><th>SVM1</th><th>SVM2</th><th>SVM3</th><th>FSVM</th></tr> </thead> <tbody> <tr><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>3</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>4</td><td>1</td><td>1</td><td>0</td><td>1</td></tr> <tr><td>5</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>6</td><td>1</td><td>1</td><td>0</td><td>1</td></tr> </tbody> </table>	N-42	SVM1	SVM2	SVM3	FSVM	1	1	1	1	1	2	1	1	1	1	3	1	1	1	1	4	1	1	0	1	5	1	1	1	1	6	1	1	0	1	<table border="1"> <thead> <tr><th>K-1</th><th>SVM1</th><th>SVM2</th><th>SVM3</th><th>FSVM</th></tr> </thead> <tbody> <tr><td>1</td><td>1</td><td>0</td><td>1</td><td>1</td></tr> <tr><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>3</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>4</td><td>1</td><td>1</td><td>0</td><td>1</td></tr> <tr><td>5</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>6</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> </tbody> </table>	K-1	SVM1	SVM2	SVM3	FSVM	1	1	0	1	1	2	1	1	1	1	3	1	1	1	1	4	1	1	0	1	5	1	1	1	1	6	1	1	1	1	<table border="1"> <thead> <tr><th>N-25</th><th>SVM1</th><th>SVM2</th><th>SVM3</th><th>FSVM</th></tr> </thead> <tbody> <tr><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>3</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>4</td><td>1</td><td>1</td><td>0</td><td>1</td></tr> <tr><td>5</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>6</td><td>1</td><td>1</td><td>0</td><td>1</td></tr> </tbody> </table>	N-25	SVM1	SVM2	SVM3	FSVM	1	1	1	1	1	2	1	1	1	1	3	1	1	1	1	4	1	1	0	1	5	1	1	1	1	6	1	1	0	1	<table border="1"> <thead> <tr><th>N-49</th><th>SVM1</th><th>SVM2</th><th>SVM3</th><th>FSVM</th></tr> </thead> <tbody> <tr><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>3</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>4</td><td>1</td><td>1</td><td>0</td><td>1</td></tr> <tr><td>5</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>6</td><td>1</td><td>1</td><td>0</td><td>1</td></tr> </tbody> </table>	N-49	SVM1	SVM2	SVM3	FSVM	1	1	1	1	1	2	1	1	1	1	3	1	1	1	1	4	1	1	0	1	5	1	1	1	1	6	1	1	0	1
N-42	SVM1	SVM2	SVM3	FSVM																																																																																																																																											
1	1	1	1	1																																																																																																																																											
2	1	1	1	1																																																																																																																																											
3	1	1	1	1																																																																																																																																											
4	1	1	0	1																																																																																																																																											
5	1	1	1	1																																																																																																																																											
6	1	1	0	1																																																																																																																																											
K-1	SVM1	SVM2	SVM3	FSVM																																																																																																																																											
1	1	0	1	1																																																																																																																																											
2	1	1	1	1																																																																																																																																											
3	1	1	1	1																																																																																																																																											
4	1	1	0	1																																																																																																																																											
5	1	1	1	1																																																																																																																																											
6	1	1	1	1																																																																																																																																											
N-25	SVM1	SVM2	SVM3	FSVM																																																																																																																																											
1	1	1	1	1																																																																																																																																											
2	1	1	1	1																																																																																																																																											
3	1	1	1	1																																																																																																																																											
4	1	1	0	1																																																																																																																																											
5	1	1	1	1																																																																																																																																											
6	1	1	0	1																																																																																																																																											
N-49	SVM1	SVM2	SVM3	FSVM																																																																																																																																											
1	1	1	1	1																																																																																																																																											
2	1	1	1	1																																																																																																																																											
3	1	1	1	1																																																																																																																																											
4	1	1	0	1																																																																																																																																											
5	1	1	1	1																																																																																																																																											
6	1	1	0	1																																																																																																																																											
<table border="1"> <thead> <tr><th>K-3</th><th>SVM1</th><th>SVM2</th><th>SVM3</th><th>FSVM</th></tr> </thead> <tbody> <tr><td>1</td><td>0</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>3</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>4</td><td>1</td><td>1</td><td>0</td><td>1</td></tr> <tr><td>5</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>6</td><td>1</td><td>1</td><td>0</td><td>1</td></tr> </tbody> </table>	K-3	SVM1	SVM2	SVM3	FSVM	1	0	1	1	1	2	1	1	1	1	3	1	1	1	1	4	1	1	0	1	5	1	1	1	1	6	1	1	0	1	<table border="1"> <thead> <tr><th>K-5</th><th>SVM1</th><th>SVM2</th><th>SVM3</th><th>FSVM</th></tr> </thead> <tbody> <tr><td>1</td><td>0</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>3</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>4</td><td>1</td><td>1</td><td>0</td><td>1</td></tr> <tr><td>5</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>6</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> </tbody> </table>	K-5	SVM1	SVM2	SVM3	FSVM	1	0	1	1	1	2	1	1	1	1	3	1	1	1	1	4	1	1	0	1	5	1	1	1	1	6	1	1	1	1																																																																								
K-3	SVM1	SVM2	SVM3	FSVM																																																																																																																																											
1	0	1	1	1																																																																																																																																											
2	1	1	1	1																																																																																																																																											
3	1	1	1	1																																																																																																																																											
4	1	1	0	1																																																																																																																																											
5	1	1	1	1																																																																																																																																											
6	1	1	0	1																																																																																																																																											
K-5	SVM1	SVM2	SVM3	FSVM																																																																																																																																											
1	0	1	1	1																																																																																																																																											
2	1	1	1	1																																																																																																																																											
3	1	1	1	1																																																																																																																																											
4	1	1	0	1																																																																																																																																											
5	1	1	1	1																																																																																																																																											
6	1	1	1	1																																																																																																																																											

Fig. 2. Classification result for PIH and control

Table 3. Parameters of SVMs

Nr	Kernel	parameters	Nr	Kernel	parameters
1	RBF	$\sigma = 1$	4	poly	$k = 1$
2	RBF	$\sigma = 10$	5	poly	$k = 2$
3	RBF	$\sigma = 100$	6	poly	$k = 3$

Table 4. Classifications result for three classes

	Number of cases	%of all cases	%of used
use of FSVM	59	61.46%	
improvement	39	40.62%	66.1%
no improvement	9	9.37%	15%
deterioration	11	11.46%	18.64%

82.5% accuracy of FSVMs was higher or equal to the average accuracy of three SVMs. Only in 17.5% cases these accuracies were lower.

5.2 Multi-classes

For simulating the proposed fuzzy SVMs the datasets from Kent Ridge Biomedical Data Set Repository is used [1]. This set consists of samples from three classes of patients: patients suffering from ALL (acute lymphocytic leukemia), therein patients suffering from ALL-T (T-cell acute lymphocytic leukemia) – 9 probes, ALL-B (B-cell acute lymphocytic leukemia) – 38 probes, and patients with AML (acute myeloblastic leukemia) – 25 probes. Like in the first experiment (for two classes) computer simulations have been performed for various configurations of conventional SVMs Table 3. For each tests, a set of all probes has been divided into two parts: testing and training set. Numerical experiments have been performed for One-against-One as well as One-against-All. In total 96 tests have been executed. Table 4. shows final result of experiments. It contains information showing how many cases of testing samples are unclassifiable using conventional SVMs and how many cases using fuzzy SVMs improved conventional classification results. Total result has been improved in 66.1% cases.

6 Conclusions

Results show the performance improvement of the FSVM over the conventional SVM. The generalization ability of fuzzy support vector machine is better in comparison with the conventional solution. Computer simulations have demonstrated that fuzzy support vector machine is also more powerful than the conventional method and can be successfully used for gene expression data analysis. This statement is confirmed by the result for two-classes as well as for multi-classes.

References

1. Li, J., Liu, H.: Kent ridge biomedical data set repository (2003), <http://sdmc.i2r.a-star.edu.serp/>
2. Cortes, C., Vapnik, V.: Support-vector networks *Machine Learning*. 20(3), 273–297 (1995)
3. Chen, X., Harrison, R., Zhang, Y.Q.: Fuzzy support vector machines for biomedical data analysis *IEEE. Granular Computing* 1(1), 131–134 (2005)
4. Inoue, T., Abe, S.: Fuzzy support vector machines for pattern classification. In: *ESANN 2002 proceedings - European Symposium on Artificial Neural Networks*, pp. 113–118 (2002)
5. Anthony, G., Gregg, H., Tshilidzi, M.: Image classification using svms: One-against-one vs one-against-all, [arXiv:0711.2914](https://arxiv.org/abs/0711.2914) (2007)
6. Abe, S., Inoue, T.: Fuzzy support vector machines for multiclass problems. In: *ESANN 2002 proceedings - European Symposium on Artificial Neural Networks*, pp. 113–118 (2002)
7. Sławska, H., Więclawek, A., Janikowska, G., Mazurek, U., Owczarek, A., Szota, J.: Expression profiles of genes encoding biogenic amine transporters and receptors in human placenta; normal versus complicated pregnancy. In: *11th Conference of the Polish Histamine Research Society - Biogenic Amines and Related Biologically Active Compounds. abstr.*, p. 7 (2006)
8. Musioł, J.: Opracować metodę analizy ekspresji genów wykorzystując elementy logiki rozmytej. MS Thesis, Silesian University of Technology, Gliwice (2007) (in Polish)

Predictive Performance of Top Differentially Expressed Genes in Microarray Gene Expression Studies

Henryk Maciejewski

Institute of Computer Engineering, Control and Robotics,
Wrocław University of Technology,
ul. Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
Henryk.Maciejewski@pwr.wroc.pl

Summary. This paper reports a comparative study demonstrating what level of predictive performance can be achieved if class prediction is attempted based on features obtained as the top most differently expressed genes from class comparison studies. Several typically used methods of gene ranking in class comparison are considered including Wilcoxon rank test, signal to noise and fold-change method. Predictive performance is estimated for a variety of feature set dimensionalities, this allows to empirically find a classification model yielding best performance for new data. This is used as a measure of predictive performance of feature vectors. Predictive performance is illustrated using publicly available microarray data sets. Results are compared with those using feature selection methods aiming to reduce feature redundancy.

1 Introduction

One of most promising application areas of microarray gene expression studies seems to be class prediction, i.e., building models for predicting classes (e.g., diseases or phenotypes) of samples based on their gene expression profiles ([2, 3]). This can open new opportunities in medical diagnosis or prediction of response to treatment, etc. Although more wide-spread use of microarrays in the clinical or regulatory practice still requires resolution of some issues related to data quality and data analysis ([4, 8]), the US FDA recently approved first microarray chip to help administer patient dosages of drugs that are metabolized differently by cytochrome P450 enzyme variant. This shows that the time for bedside application of microarrays is coming.

Building a class prediction model based on microarray data is a challenging task due to very high dimensionality of data and small number of samples available. E.g., microarray data realize dimensionality $d \sim 10^3$ to 10^4 (the number of transcripts observed in one DNA chip), while the number of samples tested is at most $n \sim 10^2$. This produces an ill-formulated problem of classification, where $d \gg n$, and requires significant dimensionality reduction, the process referred to as feature selection.

Perhaps the most widely used approach to feature selection involves ranking genes by some measure of their individual relevance to the target classes. Then

the top-ranked genes are used as features for class prediction, with the number of features selected either heuristically (as in e.g., [2]) or determined adaptively in order to control classifier complexity and avoid oversimplifying or overfitting of the model (as described and demonstrated in [9] and [6]). These simple methods of gene selection have a common shortcoming as they do not control the relationships or redundancy among features selected which can result in decreased predictive performance of features selected. This motivated development of feature selection methods analyzing correlations between pairs of genes (e.g., [12]), which aim to increase informative value of a feature set by removing redundancy. However, simple gene-ranking based feature selection seems to form the mainstream in microarray literature as it is both efficient (linear complexity in terms of the number of genes) and considered quite effective as recently pointed out by [10] who claims that "Classifiers based on rather small numbers of the top differentially expressed genes selected in sample comparison studies are usually very effective in predicting taxonomies related to future measurements".

This work aims to quantitatively measure what level of performance of class prediction for new data can be expected using features selected by several widely used gene-ranking methods. The next section explains how predictive performance for new data will be measured. Then an empirical study is shown to compare predictive performance obtained using different gene ranking methods. Results are also discussed in comparison with redundancy reduction approach reported in [12].

2 Performance of Class Prediction Using Gene Ranking Feature Selection

In this section we describe how performance of class prediction using a specific gene ranking method for feature selection can be estimated. The method described here, and developed more extensively in [7], will be used in the empirical study in section 3.

In order to estimate performance of a class prediction model built using a given gene ranking method for feature selection two major challenges have to be solved:

- the *right dimensionality* of the feature set has to be selected to avoid oversimplifying or overfitting of the model [9],
- predictive performance of the model for *new* data has to be estimated, based on test samples not used at any stage of model building (this also means that the test data should not be used for feature selection), [9, 11].

The former challenge will be approached by searching through the list of candidate feature set dimensionalities up to the number of samples, and using performance of the best model as the measure of effectiveness of the gene-ranking method applied for feature selection. In this way we avoid making a heuristic a

priori assumption on the number of features in the model (as this assumption can bias the estimated predictive performance if the model is too simple or too complex). Limiting the number of candidate features can be justified by the well known fact pertaining to binary classification that in d dimensions $d + 1$ points can be always perfectly separated by a simplest linear classifier [5]. Hence, by increasing the number of features perfect fit can be achieved for the training data, however deteriorating performance for *new* data is observed [9].

The latter challenge related to proper estimation of predictive performance of the model for new data will be approached by using *internal* cross-validation. This procedure requires that available data is repeatedly split into training and test parts, with the training part used for both feature selection and model fitting, leaving the test part solely for estimation of effectiveness of the model. Note that failing to include the feature selection stage within the cross-validation loop (referred to as *external* cross-validation) can significantly bias the estimated predictive performance for new data, as shown e.g., in [11].

The method used in the empirical study will be described formally using the following notation. Let $x_i, y_i, 1 = 1, 2, \dots, n$ denote data from the n samples tested, where $x_i \in R^d$ represents gene expressions from sample i and $y_i \in C = \{c_1, c_2\}$ denotes the known class membership associated with the sample i . (Here we consider only the binary classification; this can be extended to the multi class problem by using ANOVA based metrics for gene ranking such as the F-statistic).

Performance of a class prediction model $f : R^d \mapsto C$ can be estimated by *empirical risk* defined as

$$\frac{1}{k} \sum_{i=1}^k L(y_i, f(x_i))$$

where the *loss function* L equals 1 for $f(x_i) \neq y_i$ and 0, otherwise, and is used to punish misclassification errors. Note that empirical risk should be computed on data not used for building the model f . Empirical risk estimates the *expected prediction error (EPE)* of f defined in statistical learning theory [5].

Since in microarray studies the number of samples n is usually small, performance of the model f for new data (or *EPE*) can be estimated with leave-out-one cross-validation as:

$$EPE \approx \frac{1}{n} \sum_{i=1}^n L(y_i, f^{-i}(x_i))$$

where f^{-i} is the classifier fitted to data with the sample x_i removed [5]. As mentioned previously, feature selection will also be done with the sample x_i removed (internal cross-validation).

In section 3 we plot the *EPE* versus the dimensionality of the feature set. This gives insight into how performance of class prediction depends on the gene-ranking method used for feature selection, and allows to select the best performance model. It is performance of this model that will be used as the quality measure to rank feature selection methods.

3 Empirical Study

This empirical study illustrates what performance for class prediction can be achieved using the following gene-ranking methods for feature selection:

- Wilcoxon rank-test,
- Fold difference (used and recommended by [8]),
- Signal to noise measure.

Feature selection using Wilcoxon test ranks genes by increasing value of the p-value of the nonparametric group comparison test, performed independently for every gene. This returns the set of top differentially expressed genes between the compared classes c_1 and c_2 .

Feature selection using the fold difference measure ranks genes by decreasing value of ratio of mean expression from samples of class c_1 and c_2 . More specifically, if for a given gene, mean value of gene expression from samples of class c_1 and c_2 is denoted μ_1 and μ_2 , then the (log) fold difference measure used for gene ranking is defined as $fc = |\log(\mu_1) - \log(\mu_2)|$, which produces high values if either of the means exceeds the other.

Feature selection based on the signal to noise ranks genes by decreasing value of the measure defined as $sn = \frac{|\mu_1 - \mu_2|}{\sigma_1 + \sigma_2}$, where σ_1 and σ_2 are standard deviations of expressions of a fixed gene for samples of class c_1 and c_2 , respectively.

Performance of class prediction for these methods of feature selection is demonstrated using two publicly available microarray data sets: colon cancer [1] and leukemia [2]. Colon data contains 62 samples (40 tumor vs 22 normal) with 2000 genes. Leukemia data contains 72 samples (47 ALL vs 25 AML) with 1707 genes (original dimensionality of leukemia data 7129 was reduced by removing genes with detection level of Absent in more then 30 samples out of 72).

Figures 1 and 2 compare the estimated prediction error (*EPE*) vs. model dimensionality for Wilcoxon, signal to noise and fold change gene-ranking methods using the colon cancer and leukemia data sets, respectively. In these examples multilayer perceptron (MLP) was used (implemented as SAS `neural` procedure). In figure 3 performance of decision tree model for colon data is shown (implemented as SAS `split` procedure with entropy reduction model building criterion). We observe the following:

- Expected performance of class prediction for new data (*EPE*) tends to deteriorate for extreme (small or large) numbers of features used, realizing best *EPE* for moderate numbers of genes (around 20-30). This demonstrates the effect of over-simplifying or overfitting the model when using inappropriate number of features. The best performance model can be selected using plots as shown in Figs. 1 and 2.
- Performance of class prediction depends on the gene-ranking method used to select features. Generally, for wide range of numbers of features, Wilcoxon method yields fewer prediction errors then signal to noise or fold difference. Also, Wilcoxon method yields best class prediction ($EPE=0.16$ at 20 genes

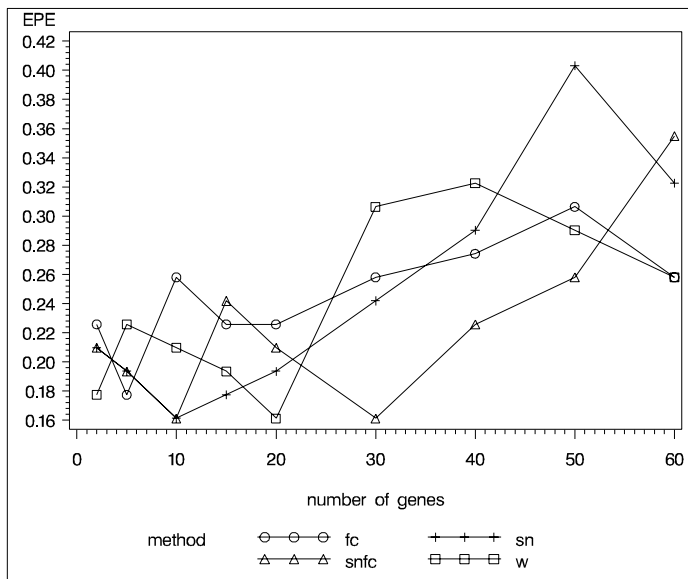


Fig. 1. Expected prediction error for colon data. Neural network model, different methods of gene selection. (Notation: w=Wilcoxon test, fc=fold difference, sn=signal to noise, snfc=signal to noise with additional fold difference criterion).

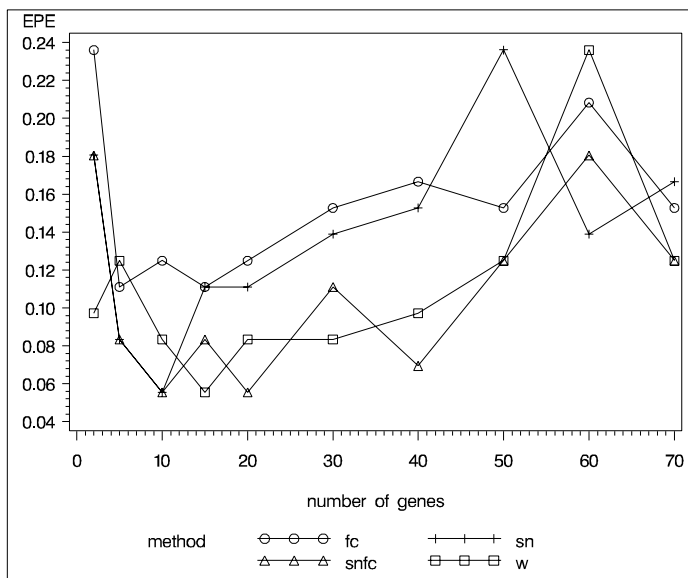


Fig. 2. Expected prediction error for leukemia data. Neural network model, different methods of gene selection. (Notation: w=Wilcoxon test, fc=fold difference, sn=signal to noise, snfc=signal to noise with additional fold difference criterion).

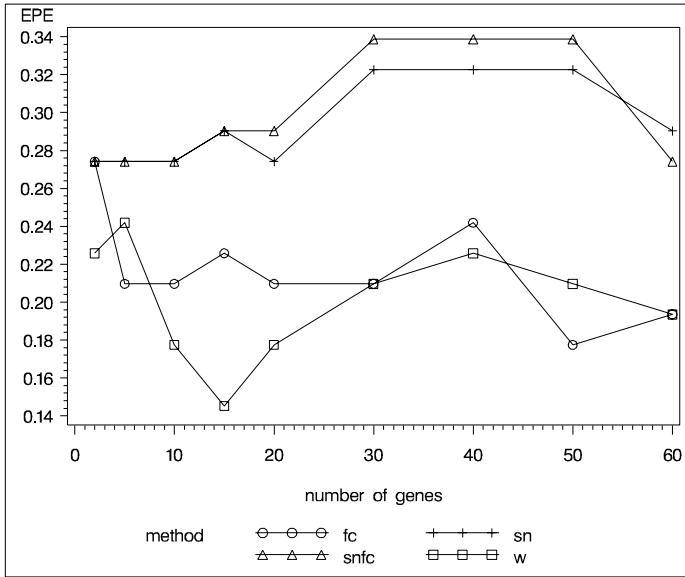


Fig. 3. Expected prediction error for colon data. Tree model, different methods of gene selection. (Notation: w=Wilcoxon test, fc=fold difference, sn=signal to noise, snfc=signal to noise with additional fold difference criterion).

for colon data; $EPE=0.056$ at 15 genes for leukemia data with the neural model; $EPE=0.145$ at 15 genes for colon data with the tree model).

- It is also interesting to notice that a combined method of feature selection (signal to noise with additional criterion of at least two-fold change in expression – shown in the figures as 'snfc') yields better performance than individual methods. (This observation holds for the neural network classifier).

Yu and Liu in [12] report results of class prediction (with tree model) using more sophisticated methods of gene selection, such as Redundancy Based Filter (RBF). Such methods aim to produce feature sets with limited redundancy and thus overcome the fundamental shortage of simple gene-ranking feature selection considered in this paper. The RBF feature selection yields the following results: $EPE=0.065$ for the colon data and $EPE=0.125$ for leukemia, each with 4 genes selected). For colon data their results are remarkably better than the best performance of Wilcoxon features (0.065 vs 0.145), however for leukemia data Wilcoxon features outperform RBF features (0.125 vs 0.056). However, since results published in [12] were estimated using *external* cross-validation (i.e., test data was taken into consideration for feature selection), we advise that their results are taken with caution, as they may be over-optimistic, as shown in [11] and [9]. The effect of using external cross validation is illustrated in fig. 4. Our observations consistently show that for small numbers of features external cross-validation leads to unrealistic estimates of performance of class prediction.

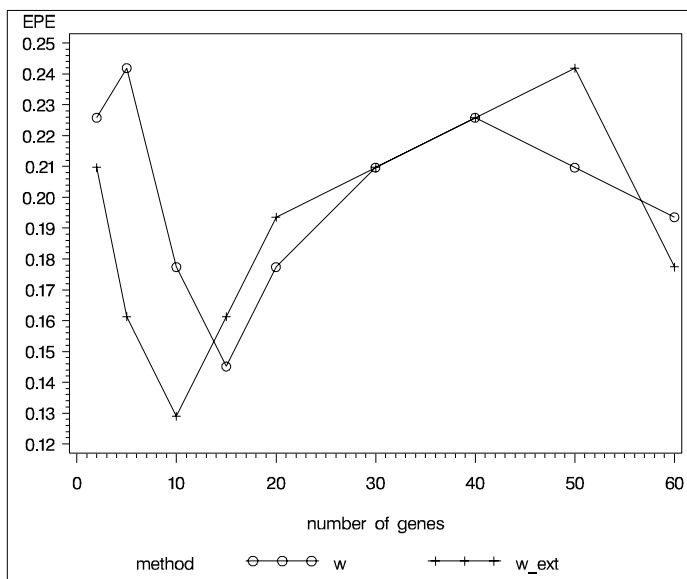


Fig. 4. Expected prediction error for colon data, measured by internal or external cross-validation. Tree model, Wilcoxon feature selection. (Notation: w=internal cross-validation, w_ext=external cross-validation).

4 Conclusions

This study demonstrates that performance of simple gene-ranking methods used for feature selection in class prediction varies considerably among such commonly used methods as Wilcoxon nonparametric test, signal to noise and fold change. Out of these three approaches, Wilcoxon method produces most informative features, i.e., allows to build the most efficient classifiers. Interestingly, the study based on leukemia data shows that Wilcoxon features outperform in terms of the classifier performance low-redundancy features obtained with a computationally more expensive RBF method. This work also suggests that simple gene-ranking feature selection should be extended to control feature redundancy, and similarly, redundancy reduction methods of feature selection should benefit from being able to adaptively control the number of features returned. This however is a field for further research.

References

1. Alon, U., et al.: Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc.Natl Acad. Sci.* 96, 6745–6750 (1999)
2. Golub, T., et al.: Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* 286, 531–537 (1999)

3. Gordon, G.J., et al.: Translation of Microarray Data into Clinically Relevant Cancer Diagnostic Tests Using Gene Expression Ratios in Lung Cancer and Mesothelioma. *Cancer Research* 62, 4963–4967 (2002)
4. Guo, L., et al.: Rat toxicogenomic study reveals analytical consistency across microarray platforms. *Nature Biotechnology* 24, 1162–1169 (2006)
5. Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning*. In: *Data Mining, Inference and Prediction*, Springer, Heidelberg (2002)
6. Maciejewski, H.: Adaptive selection of feature set dimensionality for classification of DNA microarray samples. In: *Computer recognition systems CORES 2007*. Springer *Advances in Soft Computing* (2007)
7. Maciejewski, H.: Quality of feature selection based on microarray gene expression data. *ICCS 2008* (submitted, 2008)
8. Shi, L., et al.: MAQC Consortium. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nature Biotechnology* 24, 1151–1161 (2006)
9. Markowetz, F., Spang, R.: Molecular diagnosis. Classification, Model Selection and performance evaluation, *Methods Inf. Med.* 44, 438–443 (2005)
10. Polanski, A., Kimmel, M.: *Bioinformatics*. Springer, Heidelberg (2007)
11. Simon, R., et al.: Pitfalls in the Use of DNA Microarray Data for Diagnostic and Prognostic Classification. *Journal of the National Cancer Institute* 95, 14–18 (2003)
12. Yu, L., Liu, H.: Redundancy based feature selection for microarray data. In: *Proc. of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York (2004)

A Study on Diagnostic Potential of a Computer-Assisted System for Identification of Neoplastic Urothelial Nuclei from the Bladder

A. Dulewicz, D. Piętka, and P. Jaszczak

Institute of Biocybernetics and Biomedical Engineering PAS, 02-109 Warsaw,
ul. Ks. Trojdena 4, Poland
ibib@ibib.waw.pl

Summary. The aim of the study was to test diagnostic potential of a computer-assisted system for identification of neoplastic urothelial nuclei. Presence of neoplastic urothelial nuclei in organic fluid points to neoplastic changes. The system analyzed Feulgen stained cell nuclei obtained with bladder washing technique. Image analysis was carried out by means of a digital image processing system designed by the authors. Features describing nuclei population were measured, then a multistage classifier was constructed to identify positive and negative cases. The principle of the worked out urothelial nuclei analysis on the basis of nuclei size distribution and the basic idea of the case classification were presented. The results obtained in a study of 38 new cases were compared with those obtained with earlier studies. All together 170 cases were analyzed. The results of this new study together with earlier investigated cases yielded ~60% correct classification rate in the control group, while a 86% was obtained among the cancer patients. The predictive value of the positive result of the test based on this method showed to be ~82% and the predictive value of the negative result occurred to be ~75%.

The results shown that this system may be sufficiently well developed to be used successfully in clinical practice.

1 Introduction

It is generally understood that tumors make one of the most dangerous and difficult disease from the therapeutic point of view. Bladder carcinomas are distinguished as either superficial (TTC) or muscle invasive tumors. Invasive tumors are generally associated with poor prognosis. Recurrence rate of superficial bladder tumors is high (80%) and 40% of them will progress to a muscle-invasive stage with poorer prognosis [9]. In situation when at the moment none univocal cause of tumors incidence nor a high effective therapy are found it seems that there is nothing left but early diagnosis of neoplastic changes. Process of tumor formation from the moment of action of carcinogenic factor is usually long. It is period of latency which for some tumors continue even 20 years. For the others is shorter but it is always a certain period of time, called preclinical, in which tumors are not detected. In this time symptoms are not observed but tumors can be detected. It would allowed for increase of chance for efficient therapy or prolongation of patient's life.

Urinary bladder cancer is one of the most frequent malignant neoplasm. In Poland it ranks as fourth cause of male mortality and it occur five times more often with men than women. It is diagnosed mainly in elder people: women over 60ty and men over 50ty. It causes death of more than 70% of diagnosed people in Poland (data of 1990) and it happens mainly because the illness is diagnosed in its advance stage. Early diagnose of neoplastic changes, besides increasing chance for efficient therapy, would considerably lower costs of treatment, which in case of advanced stage of invasive cancer is expensive, as it requires surgical treatment (extensive intervention of removal of an attacked organ together with lymph nodes), radiotherapy and chemotherapy.

Traditionally, the diagnosis of urinary bladder cancer is carried out by urologists (cytology, clinical observation) and histopathologists. Cytological visual examination of urine provides many false negative results especially for cases at an early grade of malignancy whereas histological examination is an invasive test. To carry out histological examination patients undergo a transurethral resection of the tumor. This intervention is associated with a number of disadvantages due to its invasive nature. It is neither entirely safe nor free of complications. It is also traumatic for patients often causing inflammation, dysuria, hematuria and difficult for the medical staff. Moreover, it is not always perfectly conclusive as you can not inspect the whole surface of the bladder and may need to be repeated several times during therapy and later in the case of relapses. An additional problem is lack of qualified cytopathologists and histopathologists at many medical centers.

Other methods used in urology such as urography, ultrasonography, computer tomography CT, magnetic resonance MRI, scintigraphy, positron emission tomography PET are applied to detect enlarged changes (bigger than 10mm) and possible metastases and are not pertinent in detecting early bladder tumors. Biochemical tests being nowadays widely tested are not accepted so far as a gold standard for detecting bladder cancer as level of biomarkers can be raised with healthy people and can not be raised with cancer patients especially in early stages of disease.

Urinary bladder tumors originate from lining and glandular epithelium of urinary bladder wall or from other cells of bladder wall (they make more than 90% of bladder tumors) (Fig. 1).

Tumor cells exfoliate into urine. Exfoliation concerns both normal and cancer cells. They can be recovered from urine or obtained with "bladder washing" technique which consist in washing out bladder with physiologic saline. Then they can be subjected to analysis after being accordingly technically prepared. Classification of urinary bladder tumors grades and stages is based on changes taking place in tumor cells and changes in the architecture of a urine bladder wall. On this basis a stage of the tumor progression and grade are stated. The G1 grade is the most early one, G2 is medium and G3 is the most advanced one. The most important task for a diagnosis method is to recognise grade G1. Another applied classification of malignancy is : "low grade" for G1 grade cases and "high grade" malignancy for G2 and G3 grade cases.

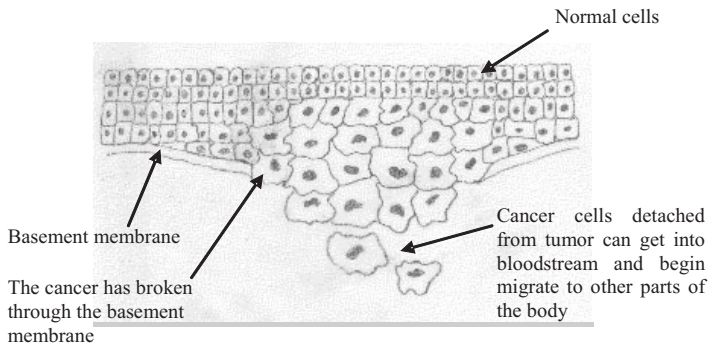


Fig. 1. Cancer cells detach from the tumor and begin migrate with blood stream to other parts of the body

As it was above mentioned cystoscopic inspection is an invasive method. Other applied methods such as cytology of urinary sediment and ultrasonography of urinary bladder are not precise methods especially in cases of small urinary bladder tumors. Therefore, working out a computer standards for recognizing neoplastic changes in bladder sediment cells being in different grades of malignancy would considerably facilitate for control examinations.

2 Materials and Methods

Cells for analysis in our study were obtained with bladder washing technique. Then material was concentrated, specially prepared to prefix the cells. Next the cells were settled on glass slides, dried, post fixed and stained (Feulgen staining). Feulgen staining was choose in order to get specimens with stained nuclei without cytoplasm. There were several reasons to analyze not the whole cells but nuclei only:

- cell nuclei provide most information for cancer detection,
- analyzing nuclei it was possible to avoid the problem of double extraction (extraction whole cells from the image background and extraction nuclei from the cytoplasm),
- avoiding sticking, deformed and overlying objects.

The specimens were observed under an optical microscope (Nikon Optiphot 2) at an objective magnification of 10:1 and were automatically scanned by means of our own software SSU (Stage Scan Utility) (Fig. 2). Observed images were transmitted to the computer system. Then, a full cycle of processing and analysis was performed by DIPS (Digital Image Processing System) software worked out by authors. DIPS software facilitates correction of the image background and normalization, which are essential conditions for correct comparison of objects

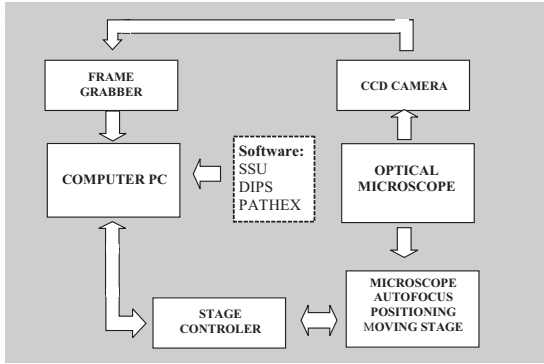


Fig. 2. The general view of the system configuration

derived from different images. The general view of the system configuration is presented in Fig. 2.

There are many image processing systems available commercially but usually they are not useful for professional medical applications. That is because they provide a lot of tools trying to be universal ones for solving all possible image processing problems but when one try to use them one will quickly find out that the software package dose not suit for an actual task. Therefore, we have designed our own software package DIPS which was a result of several years of experience and collaboration with medical and computer scientists. The basic ideas on which the project was established are following:

- almost every real research task requires special picture processing methods and algorithms depending on the nature and scope of investigations. Therefore, procedures available for the user should be specialised for strictly defined task. Their internal complexity should not be visible for him,
- flexibility and universality should be achieved by opening the system for user designed software extensions.

In 144 images registered for every one specimen 3000-18000 nuclei were registered. Usually each single image contained ~ 30 to 200 nuclei . The size of the recorded images was: 480×640 pixels (8 bits). For working out a classification rule data were obtained from cytological specimens of 56 persons. Among them 20 had no cancer (control), 19 were diagnosed as having bladder cancer of a "low grade" and 17 as having cancer of a "high grade" histopathological malignancy. Then the system was tested on a new clinical material of 114 cases to check its usefulness in clinical practice. The specimens were prepared and diagnosed in Holland at first by Department of Urology at University Hospital of Nijmegen and later by Leiden Pathological Laboratory.

2.1 Image Analysis

The DIPS's procedures were specially designed and implemented in DIPS software to meet this kind of input images (urinary bladder specimen images). The way of image **background correction** consisted in this software in creating of a two dimmetional correction matrix and using it later for "pixel" multiplication with images of the specimen. The correction matrix was created on the basis of 10 images of "empty field" derived from different places of the analysed specimen. Each image should contain a small number of pixels of maximum light intensity (it could be achieved by regulation of microscope lighting) in order to normalize conditions of lighting. The result correction image (an averaged image of empty field) was obtained on the basis of pixels light intensity of the all composed images. The light intensity values of the same coordinates pixels were taken, two extremes values were neglected and the average value was calculated out of the rest. Thus the correction image free from individual features such as scratches, spots, etc. was obtained.

A value of the correction matrix elements (coefficients) was calculated according to following formula:

$$M(i, j) = \max/O(i, j),$$

where $M(i, j)$ – value of a correction matrix coefficient for (i, j) coordinates of the correction matrix,

\max – maximum value of light intensity in the image (saturation),

$O(i, j)$ – calculated average value of a correction image in (i, j) element.

In elements where a correction image had maximal values, correction coefficients (elements of correction matrix) were equal 1, in all the others had values bigger than 1.

Influence of intensity light changes of a specimen had a multiplicative character. It means that percentage change of light intensity was the same for all pixels of the image. Images which were to be compare should had the same light intensity of the backgrounds.

The normalization brightened input images in that way that the most frequent pixels intensity in the background gained saturation (the maximum of light intensity). It was realized by the help of multiplication of all the image pixels by a coefficient being found for each image on the basis of its light intensity histogram. It was calculated as quotient of the saturation and maximum value of the image histogram.

$$N = 255/J\max,$$

where N – normalization coefficient, $J\max$ – intensity level found as maximum value of the image histogram.

Fig. 3. shows histograms of light distribution in the image before normalization. The maximum value of the histogram light distribution was found as mean value of its highest values.

Fig. 4 shows histograms of light distribution in the image before and after normalization.

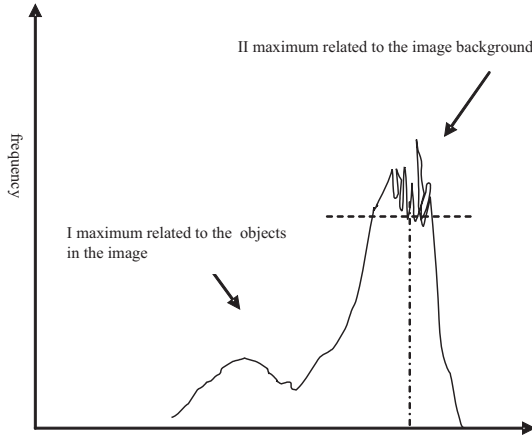


Fig. 3. The histogram of the image light distribution before the normalization process

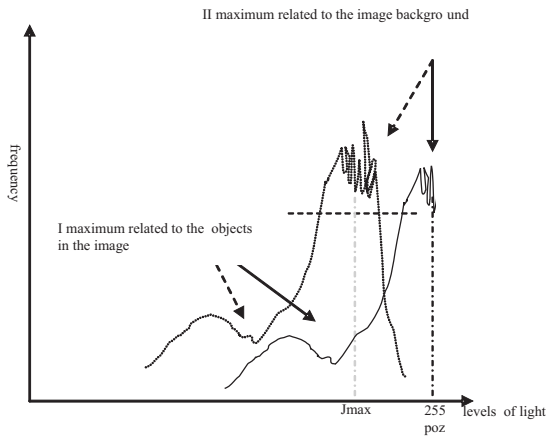


Fig. 4. Schematic presentation of light distribution histograms in an image before and after performing the normalization procedure. (The first and second maximum of light distribution marked). - dotted line shows histogram of light distribution in the image before normalization, - continuous line shows histogram of light distribution in the image after normalization.

The next processing stages of image analysis were:

- object extraction,
- measurement of selected features,
- classification of cases.

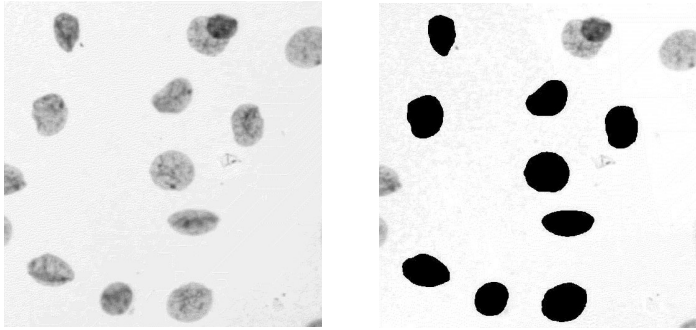


Fig. 5. The example of object extraction (microscope magnification $\times 60$)

The algorithm of **objects extraction** was also specially designed for these type of images and had been based on the image histogram of light distribution. The minimum of the histogram corresponded to a light intensity level where there was a passage between background and objects light intensity levels (Fig. 3, Fig. 4). The value of this minimum was used for performing the thresholding operation.

Fig. 5. shows the captured microscopic image and the same image with extracted objects.

Subsequent stage of processing to object extractions is **parameters measurements** of selected features. The choice of appropriate features is essential for the task as it should describe a class of objects as univocal as possible and at the same time be essential from the discrimination point of view. One never knows in advance which set of features will discriminate the best analysed classes of objects giving the smallest classification error. In this approach we based on penetrating observation of data and knowledge gathered by experienced medical staff.

Neoplastic cells (Fig. 7) differ from normal cells (Fig. 6) in many ways as they have:

- different structure of nuclei (Fig. 7),

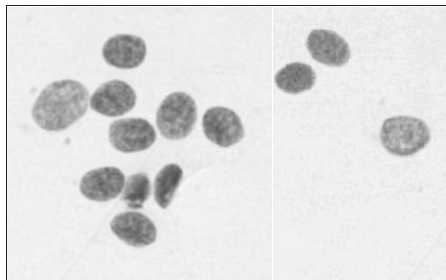


Fig. 6. The example of normal cells(microscope magnification x 60)

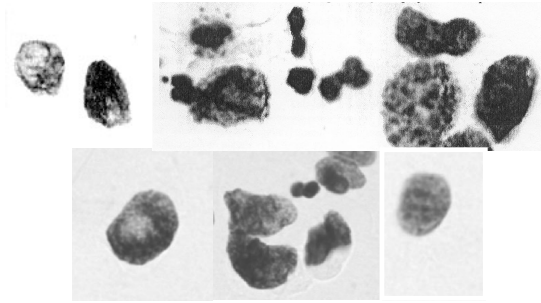


Fig. 7. The example of neoplastic nuclei. Visible bigger size, shape and structure of neoplastic nuclei (microscope magnification x 60).

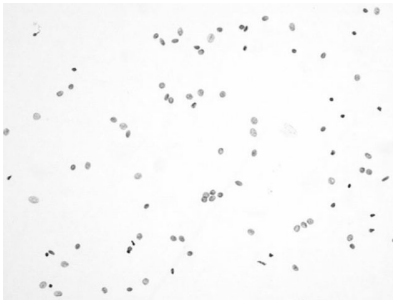


Fig. 8. The example of normal nuclei. Visible less number of nuclei in the image of normal case (microscope magnification x 10).

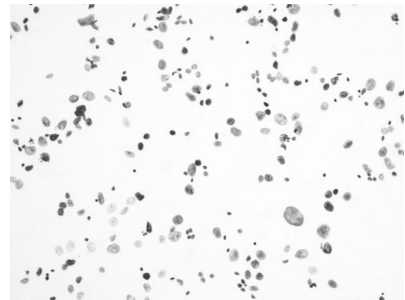


Fig. 9. The example of neoplastic nuclei. Visible great number of nuclei in the image of cancer case (microscope magnification x 10).

- increased ratio of nucleus to cytoplasm,
- bigger nuclei than in normal cells (Fig. 7),
- irregular shapes of nuclei (Fig. 7),
- nuclei contain more chromatin (Fig. 7),
- nuclei have enlarged nucleoli (Fig. 7),
- sometimes bigger number of nucleoli (Fig. 7),
- more of them exfoliate from the bladder wall (Fig. 9),
- nuclei form irregular clusters (Fig. 10).

Characteristic features of neoplastic nuclei related to their shape and inner structure was used in our other project based on morphological analysis of nuclei with a very good result but not possible so far to be used in practice as analysis lasted too long time (hours). That was because microscopic magnification of 60 times was applied.

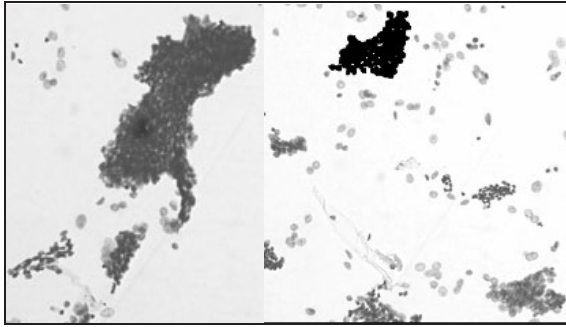


Fig. 10. Two examples of noplastic nuclei clusters in the images of cancer cases. Visible big clusters of irregular shapes.

2.2 Features Selection

In this project we concentrated on those changes occurring in neoplastic process which undergo in nuclei and concern nuclei size distribution. The general observations were:

- an increased number of bigger nuclei in images of the malignant cases (Fig. 8, Fig. 9),
- numerous exfoliation of urothelial cancer cells than normal cells (Fig. 9), What means, that density of nuclei could be a good feature for discriminating between control and malignancy.
- relatively big differences in size and structure of nuclei clusters in images of the malignant cases (Fig. 10),

Analysis of nuclei size histograms of control, "low grade" and "high grade" malignancy showed the biggest differentiation between those groups in the range of 5 - 150 pixels, what corresponds to the range of $3,6\mu\text{m}^2$ - $108\mu\text{m}^2$. Fig. 11 shows the histogram of nuclei size distribution in four ranges of nuclei size.

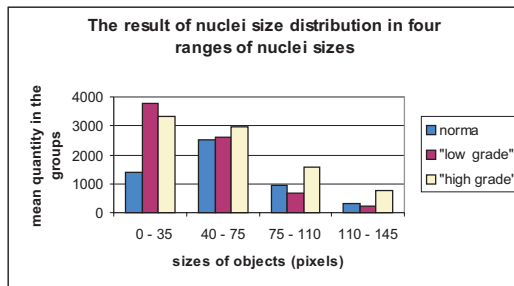


Fig. 11. The histogram of nuclei size distribution in four ranges of nuclei size

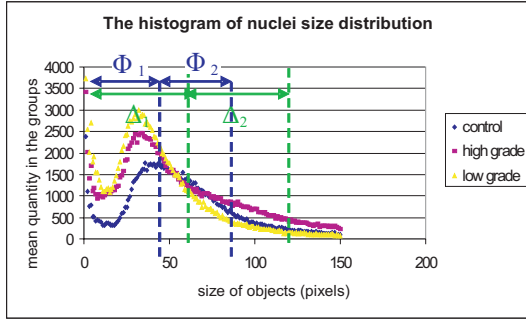


Fig. 12. The mean nuclei size distributions for control, low grade and high grade groups of cases

Finally, there were 6 parameters defined on the basis of histogram of nuclei size distribution. The list of them is presented below:

- Feature Δ , defined as a ratio of nuclei number in the range of 27-60 pixels size to number of nuclei in the range of 61-120 pixels size (Fig. 12),

$$\Delta = \frac{\Delta_1}{\Delta_2}$$

- Feature Φ , defined as a ratio of nuclei number in the range 27- 40 pixels size to number of nuclei in the range of 41 - 80 pixels size (Fig. 12),

$$\Phi = \frac{\Phi_1}{\Phi_2}$$

in both cases, if the calculated result was <1 , it indicated control, and if was ≥ 1 , it indicated malignancy.

- Feature θ , as a measure of the rate of interest of number of objects in the interval of object size: 121–500 pixels, calculated in proportion to the total number of objects in a specimen.
- Feature RB10S10, as a measure of the ratio of the number of nuclei clusters - n_k to the global number of single nuclei and granulocytes in a specimen - n_j .

$$\frac{n_k}{n_j}$$

if ratio: $RB10S10 \leq 0,04$, then it pointed out to control otherwise it denoted malignancy.

- Feature X - was a measure of the difference between low grade and high grade malignancy.

$$X = \frac{\text{number of nuclei in the range "a"}}{\text{number of nuclei in the range "b"}}$$

- where ranges: a and b concern sizes of nuclei measured in pixels Fig. 13.

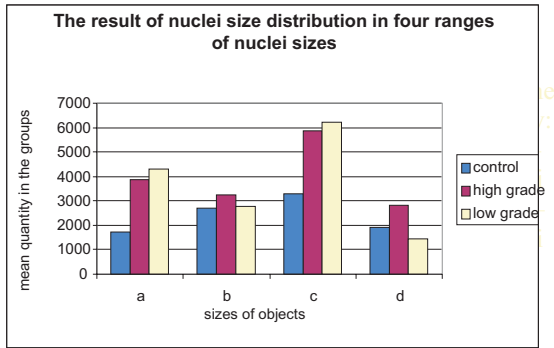


Fig. 13. The histogram of nuclei size distribution in four ranges of nuclei size, where chosen ranges: a: 1–39 pixels, b: 40–80 pixels, c: 1–59 pixels, d: 60–120 pixels

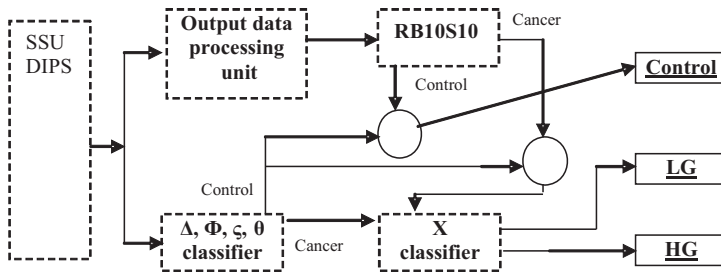


Fig. 14. The algorithm of classification, where: $\Delta, \Phi, \theta, \zeta, X$ – parameters defined as ratio of the number of objects in five different intervals of objects sizes, RB10S10 – ratio of the number of nuclei clusters to the global number of single objects (nuclei and granulocytes) in the specimen, Control – denoted norm (no cancer), LG – Low grade of histological malignancy, HG – High grade of histological malignancy

On the basis of these features multistage classifier was designed and implemented as a software package: PathEx. The Fig. 14 presents the scheme of it.

The time of one patient data analysis was 17 minutes.

3 Results

There were three stages of checking the quality of the designed system. **The first stage** was reclassification what was getting the results of the computer system classification on the data which had been used as training data.

The data were obtained from 56 patients:

- 20 had no cancer (control),
- 19 were diagnosed as having bladder cancer of "low grade",

- 17 as having cancer of "high grade" histological malignancy.

The result was: **90%** correct classification for the control group, **73,7%** correct classification for the low grade group, **88,2%** correct classification for the high grade group. It meant 10% of false positive diagnosis and $\sim 19\%$ false negative diagnosis.

The next stage of checking the quality of the designed system was carried on 76 new cases which were not used for training.

There were

- 28 patients with no cancer (control),
- 20 patients were diagnosed as having bladder cancer of "low grade",
- 28 patients were diagnosed as having cancer of "high grade" histological malignancy.

The result was: **68%** correct classification for the control group, **80%** correct classification for the low grade group, **89%** correct classification for the high grade group.

What means that the **sensitivity** of the method was = **85%**, and specificity of the method was = **68%**.

The next third stage of testing the quality of the designed system was carried on 38 new cases.

There were

- 13 patients with no cancer (control),
- 8 patients were diagnosed as having bladder cancer of "low grade",
- 14 patients were diagnosed as having cancer of "high grade" histological malignancy,
- 3 patients were diagnosed as having benign neoplasm.

The Table 1 shows the results of computer discrimination between control and malignancy.

The result of case classification of respective grades of malignancy (low and high grade), benign neoplasm and norm are shown in Table 2.

In the third testing test the result of correct classification of malignancy was better as sensitivity of the method was 88% but there was worse differentiation between low and high grade malignancy, moreover benign cases were recognized as low grade cases but the system had not been earlier thought to recognize benign cases. Another thing, was low accuracy of recognizing control cases in this group.

Table 1. The result of case classification of the third tested set of cases between "control and malignancy"

Cytopathological Classification:	Control	Neoplazm
Number of cases:	13	25
% correct computer classifications	46%	88%

Table 2. The result of case classification to respective grades of malignancy (low and high grade), benign neoplasm and norm

Actual group	No. of cases	Predicted group membership			
		Control	LG	HG	Benigen
Control	13	6	6	1	0
Low Grade	8	2	6	0	0
High Grade	14	1	13	0	0
Benigen	3	0	3	0	0

The general result of all the tested cases was:

- **sensitivity** of the method = **86.5%**,
- **specificity** of the method = **57%**.

Sensitivity and specificity show how properly a test discriminate cancer patients from control. Knowing the result of a test it is possible to determine what is the probability of a disease in case of positive result of the test. The answer for this question gives so called *predictive value* of the test, which is a measure of the ratio of really positive cases (TP) to a sum of really positive cases (TP) and false positive cases (FP). This value is also called *posttest probability* and determine probability of a disease in case of positive result of the test.

$$predictive\ value\ of\ the\ positive\ test = \frac{TP}{TP + FP}$$

For the tested method this value was equal **78%**.

Likewise knowing the result of the negative test it is possible to determine what is the credibility of this result. A *predictive value of negative test* is calculated as a measure of the ratio of really negative cases (TN) to a sum of really negative cases (TN) and false negative cases.

$$predictive\ value\ of\ negative\ test = \frac{TN}{TN + FN}$$

For the tested method this value was equal **81%**.

A diagnostic usefulness of a test in practical use depends to a great extend on purpose that it is to serve: screening examination or ascertaining of a disease. This test considering its certain degree of invasiveness should be applied in clinics not for screening examinations. Therefore, modification of thresholding value in order to increase its sensitivity can still improve parameters of this test.

In Fig. 15 a graph of ROC presents relation between sensitivity and specificity of this method. Such a relation shows how the worked out method presents against a background of graphs considered as the excellent, good and worthless results. The points marked with a star points out to the results of the worked out method.

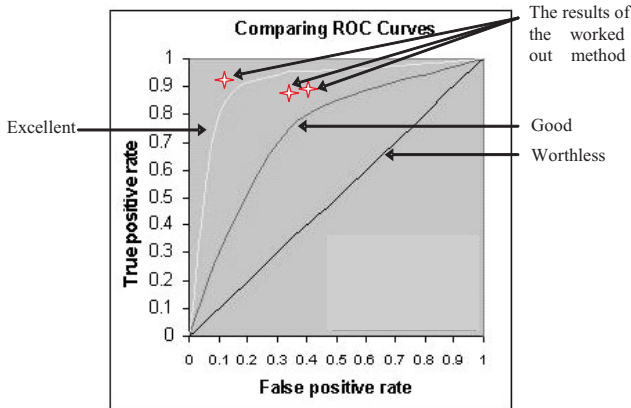


Fig. 15. The curves of ROC present relations between sensitivity and specificity of methods

4 Conclusions

The results of a pilot study carried out on 170 smears were derived from 170 patients. Among them there were 61 of the control group, 106 cancer patients and 3 benign cases. As the end result of testing the $\sim 60\%$ correct classification in the control group and 86% correct classification of cancer patients were achieved. The predictive value of the positive test was $\sim 82\%$ and the predictive value of the negative test was $\sim 75\%$.

A computer-aided system based on analysis of histograms of nuclei and their clusters size distribution in microscopic specimens of Feulgen stained urine bladder nucleated cells has shown that it could be helpful in clinical practice greatly contributing to a rapid confirmation of the diagnosis made by the physicians being an additional diagnostic tool.

The method is only a little invasive and enables diagnosing bladder cancer in early stadium and during its progression or involution in the course of therapy in a quick and easy way.

References

1. Dulewicz, A., Piętka, D., Jaszczak, P., Nechay, A., Sawicki, W., Pykało, R., Koźmińska, E., Borkowski, A.: Computer identification of neoplastic urothelial nuclei from the bladder. *Analytical and Quantitative Cytology and Histology* 23(5), 321–329 (2001)
2. Boon, M.E., Drijver, J.S.: Routine cytological staining techniques. *Theoretical Background and Practice*. Macmillan Education Ltd., London (1986)
3. Dulewicz, A., Piętka, D., Jaszczak, P., Nechay, A.: Selective acquisition of images in the process of automatic scanning of microscopic slides. In: *Proc. World Congress on Medical Physics and Biomedical Engineering, Chicago, TU–A318–08* (2000)

4. Dulewicz, A., Piętka, D., Jaszczak, P.: A method of image information selection in the process of automatic scanning of microscopic specimens. *Journal of Applied Computer Science* 11(2) (2003)
5. Dulewicz, A., Piętka, D., Jaszczak, P.: Value of digital analysis in research and diagnosis of urine bladder cancer. *Advances in Soft Computing 45: Computer Recognition Systems 2*, 45(2) (2007)
6. Dulewicz, A., Piętka, D., Jaszczak, P.: Trial of practical computer analysis of urothelial nuclei for cancer detection. In: *Progress in bladder cancer research*, pp. 173–190. Nova Biomedical Books, New York (2005)
7. Kawiak, J., Zabala, M.: *Seminarium z cytofizjologii*. Wydawnictwo Medyczne Urban & Partner, Wrocław (2006)
8. Kurzyński, M.: *Rozpoznawanie obrazów*. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław, str.12–45, 58–102, 143–216 (1997)
9. Lascomb, I., Fauconnet, S., Chabannes, E., Bittard, H.: *Angiogenesis and bladder cancer role of vascular endothelial growth factor*. Progress in bladder cancer research, Nova Biomedical Books, New York (2005)
10. Piętka, D., Dulewicz, A., Jaszczak, P.: Pathology explorer (PathEx) a computer-aided system for urinary bladder cancer detection. In: *XIII Scientific Conference Biocybernetics and Biomedical Engineering, Gdańsk, CD-ROM Proceedings, SessionXII-2* (2003)
11. Russ, J.C.: *The Image Processing handbook*. CRC Press, Boca Raton, Ann Arbor, London, Tokyo (1995)
12. Tadeusiewicz, R., Izworski, A., Majewski, J.: *Biometria*. Wyd. AGH, Kraków (1993)
13. Zieliński, K.W., Strzelecki, M.: *Computer analysis of biomedical image*. Wydawnictwo Naukowe PWN, Warszawa-Lódź (2002)
14. Zieliński, J., Leńko, J.: *Urologia. TomII, Onkologia urologiczna*. P.Z.W.L., Warszawa (1993)

Control of Hand Bioprosthesis Via Sequential Recognition of Patient's Intent Using Combination of Fuzzy Sets and Dempster-Shafer Theory

Marek Kurzynski¹, Andrzej Wolczowski², and Mariusz Topolski¹

¹ Wroclaw University of Technology, Faculty of Electronics, Chair of Systems and Computer Networks, Wroclaw, Poland

marek.kurzynski@pwr.wroc.pl

² Wroclaw University of Technology, Faculty of Electronics, Institute of Robotics and Informatics, Wroclaw, Poland

andrzej.wolczowski@pwr.wroc.pl

Summary. The paper presents a concept of bioprosthesis control via recognition of user intent on the basis of myopotentials acquired from his body. The contextual recognition of elementary actions is considered and recognition algorithm based on fuzzy inference system with mathematical evidence model is described. Proposed method was experimentally tested on the real data dealing with grasping of different objects limited to the sequences of seven steps of elementary actions.

1 Introduction

The activity of human organism is accompanied by physical quantities variation which can be registered with appropriate measuring instruments. Some of such biosignals can be applied to control the work of technical devices. This allows for new possibilities of using these devices by enabling a close integration of a machine and a living organism into one being. The only human body signals, which can be used for control purposes, are those which can be created and sent intentionally. Electrical potentials accompanying skeleton muscles' activity belong to this type of signals. They are called myopotential or electromyographic (EMG) signals.

Contemporary hand prostheses are usually based on myoelectric control. Such control takes advantage of the fact that after a hand amputation great majority of the muscles that generate finger motion is left in the stump. The activity of these muscles still depends on the patient will, so biosignals that occur during it, can be used to control prosthesis motion. In the simplest case these signals can be detected (in a non-invasive way) on the surface of the skin using electrodes located above the examined muscles. That kind of measurement is called the surface electromyography (EMG) [1].

The EMG signal obtained in such a way is the sum of electrical phenomena taking place in the cells of the working muscles. Its form depends on the level

of excitement and the spatial localization of the muscles, and that is how it identifies the type of the performed movement. This relation is the basis of the bio-prosthesis decision control. The patient tenses the muscles of the stump according to a prosthesis movement intention. The information about the type of muscles activity included in the EMG signals can be identified through adequate signal analysis (classification) [1, 4, 7, 8, 9, 10, 11].

A large repertoire of motion actions demands a large number of recognized classes of EMG signals. Furthermore, since the signal acquisition (especially performed in a non-invasive way, using electrodes located on the surface of the skin of the stump) is accompanied by numerous disturbances the reliable recognizing of the signals (especially when the number of possible classes is great) is a difficult problem. Reliability of EMG signal recognition (and thus the correctness of taking decisions by the system) is here the key issue as the prosthesis cannot perform any action inconsistent with human intent.

The paper presents the concept of a bio-prosthesis control system which consists in recognition of a prosthesis user's intention based on algorithm with fuzzy inference and mathematical evidence model. The paper arrangement is as follows. Chapter 2 includes the concept of prosthesis control system based on the recognition of patient's intent. Chapter 3 describes recognition algorithm and chapter 4 in turn presents a specific example of the described concept and its practical application for the control of dexterous hand bio-prosthesis.

2 Control System of Bio-prosthesis

In the considered control concept we assume that each prosthesis operation (irrespective of prosthesis type) consists of specific sequence of elementary actions, and the patient intention means his will to perform a specific elementary action.

Thus prosthesis control is a discrete process where at the n -th stage ($n = 1, 2, \dots, N$) occurs successively:

- the measurement of EMG signal parameters x_n , ($x_n \in \mathcal{X} \subseteq \mathcal{R}^d$), that represent patient's will j_n ($j_n \in \mathcal{M} = \{1, 2, \dots, M\}$) (the intention to take a particular action),
- the recognition of this intention (the result of recognition at the n -th stage will be denoted by $i_n \in \mathcal{M}$),
- the realisation of an elementary action $a_n \in \mathcal{A}$, uniquely defined as a recognized intention.

This means that there is M number of elementary actions $\mathcal{A} = \{a^{(1)}, a^{(2)}, \dots, a^{(M)}\}$ (an exemplary meaning of elementary actions in relation to a dexterous hand prosthesis is defined in chapter 4). The assumed character of control decisions (performing an elementary action) means that the task of bioprosthesis control is reduced to the recognition of patient's intent in successive stages on the basis of available measurement information, thus the determination of the recognition algorithm is equivalent to the determination of prosthesis control algorithm.

For the purpose of determining patient’s intent recognition algorithm, we will apply the concept of the so-called sequence recognition. The essence of sequence recognition in relation to the issue we are examining is the assumption that the intention at a given stage depends on earlier intentions. This assumption seems relevant since particular elementary actions of a prosthesis must compose a defined logical entity. This means that not all sequences of elementary actions are acceptable, only those which contribute to the activities which can be performed by a prosthesis. Examples of such actions (sequences of elementary actions) are presented in chapter 4.

Since the patient’s current intention depends on history, generally the decision (recognition) algorithm must take into account the whole sequence of the preceding feature values (parameters of EMG signal), $\bar{x}_n = (x_1, x_2, \dots, x_n)$. It must be stressed, however, that sometimes it may be difficult to include all the available data, especially for bigger n . In such cases we have to allow various simplifications (e.g. make allowance for only several recent values in the \bar{x}_n vectors), or compromises (e.g. substituting the whole activity history segment that spreads as far back as the k -th instant, i.e. the \bar{x}_k values, with data processed in the form of a decision established at that instant, say i_k)[3].

Apart from the data measured for a specific patient we need some more general information to take a valid recognition decision, namely the *a priori* information (knowledge) concerning the general associations that hold between decisions (patient’s intentions) and features (EMG signal parameters). This knowledge may have multifarious forms and various origin. From now on we assume that it has the form of a so-called training set, which - in the considered decision problem - consists of training sequences:

$$S = \{S_1, S_2, \dots, S_m\}. \tag{1}$$

A single sequence:

$$S_k = \{(x_{1k}, j_{1k}), (x_{2k}, j_{2k}), \dots, (x_{Nk}, j_{Nk})\} \tag{2}$$

denotes a single-patient sequence of prosthesis activity that comprises N EMG signal observation instants, and the patient’s intentions.

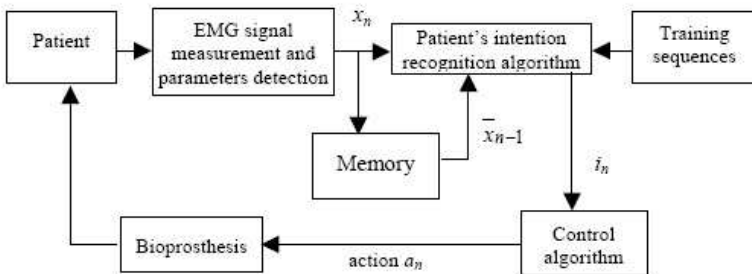


Fig. 1. System of bioprosthesis control via sequential recognition of patient’s intentions

An analysis of the sequential diagnosis task implies that, when considered in its most general form, the explored decision algorithm can in the n -th step make use of the whole available measurement data, as well as the knowledge included in the training set. In consequence, the algorithm is of the following form:

$$\psi_n(\bar{x}_n, S) = i_n. \tag{3}$$

Fig. 1 shows the block diagram of the dynamic process of prosthesis control

3 Algorithm of Sequential Recognition

With the usage of Dempster-Shafer theory it is possible to describe interactions between consecutive classes as a transition matrix:

$$\mathbf{P}^{(j_{n-1})} = \begin{bmatrix} m^{(j_{n-1}=1)}(\{1\}) & m^{(j_{n-1}=1)}(\{2\}) & \dots & m^{(j_{n-1}=1)}(\{M\}) \\ m^{(j_{n-1}=2)}(\{1\}) & m^{(j_{n-1}=2)}(\{2\}) & \dots & m^{(j_{n-1}=2)}(\{M\}) \\ \vdots & \vdots & \ddots & \vdots \\ m^{(j_{n-1}=M)}(\{1\}) & m^{(j_{n-1}=M)}(\{2\}) & \dots & m^{(j_{n-1}=M)}(\{M\}) \end{bmatrix}, \tag{4}$$

where m is the basic probability assignment function. Values in rows of the matrix (4) sum to one. It is assumed that available knowledge is available in the learning set (1).

Learning set S is used for creating fuzzy rule base system of the IF-THEN type. The conclusion of k -th rule is a discrete fuzzy set

$$Y^{(k)} = \{1/m^{(j_{n-1})}(\{1\}), 2/m^{(j_{n-1})}(\{2\}), \dots, M/m^{(j_{n-1})}(\{M\})\}, \tag{5}$$

in which values of membership function are replaced with values of basic probability assignment function, fulfilling the orthogonality condition [13]:

$$\begin{aligned} \text{a) } & \sum_{Y^{(k)} \subseteq \mathcal{M}} m^{(j_{n-1})}(Y^{(k)}) = 1, \\ \text{b) } & \forall j_{n-1} \quad m^{(j_{n-1})}(\emptyset) = 0 \end{aligned} \tag{6}$$

where

$$m^{(j_{n-1})}(Y^{(k)}) = \begin{cases} m^{(j_{n-1})}(\{1\}) & j = 1 \\ m^{(j_{n-1})}(\{2\}) & j = 2 \\ \vdots & \\ m^{(j_{n-1})}(\{M\}) & j = M \end{cases}. \tag{7}$$

During the inference process all active rules are aggregated using the Dempster-Shafer combination rule, viz:

$$m^{(j_{n-1})}(Y^*) = m^{(j_{n-1})}(Y^{(k)}) \oplus m^{(j_{n-1})}(Y^{(k-1)}) = \tag{8}$$

$$= \frac{\sum_{Y^{(k)} \cap Y^{(k-1)} = Y^*} m^{(j_{n-1})}(Y^{(k)}) \cdot m^{(j_{n-1})}(Y^{(k-1)})}{1 - \sum_{Y^{(k)} \cap Y^{(k-1)} = \emptyset} m^{(j_{n-1})}(Y^{(k)}) \cdot m^{(j_{n-1})}(Y^{(k-1)})}. \tag{9}$$

The order of joining of active rules is strictly restricted. Combination process starts with the rule with the highest activation coefficient and ends with the rule with the lowest activation coefficient. The activation coefficient of rule is calculated as a product of values of premises membership functions.

Further combination of other active rules is performed only if the non-contradiction condition is met:

$$\sum_{Y^{(k)} \cap Y^{(k-1)} = \emptyset} m^{(j_{n-1})}(Y^{(k)}) \cdot m^{(j_{n-1})}(Y^{(k-1)}) < \alpha_F, \tag{10}$$

where $\alpha_F \in < 0, 1 >$ is the contradiction coefficient calculated empirically. This can be done by discretizing its values e.g. with the step 0.01 in an experimental manner. The value of α_F for which the classifier achieves the best performance should be treated as the definitive one.

The next step is to calculate the belief function [13]

$$Bel^{(j_{n-1})}(Y) = \sum_{Y^* \cap Y} m^{(j_{n-1})}(Y^*), \tag{11}$$

and hence we simply get the following sequential classifier:

$$\Psi_n(x_n, S) = i_n \quad \text{if} \quad Bel^{(j_{n-1})}(\{i_n\}) = \max_{r \in \mathcal{M}} Bel^{(j_{n-1})}(\{r\}). \tag{12}$$

4 Experimental Investigations

The presented above sequential recognition algorithm has been tested experimentally on real data, in an example of the task of controlling a hand prosthesis model. In the considered example, seven steps can be distinguished in the process of grasping with a hand [5]:

- a_0 – rest position (starting point for the grasp preparation; the fingers stay at rest - half-closed arrangement, are motionless and passive),
- a_1 – grasp preparation (the fingers are "opening" or "closing", taking the posture depending on the shape of the object that is observed visually and the knowledge K about the method of grasping it, with the velocity proportional to the velocity of the intended arm movement),
- a_2 – grasp closing (precedes the grasp - the fingers move with the velocity resulting from the observation and knowledge about the object behaviour and the movement velocity during the previous stage, taking the posture depending on grasping object),

- a_3 – grabbing (the fingers squeeze the object with a force dependent on the knowledge and observed behaviour of the object and proportionally to the arm motion velocity in the grasp preparation phase),
- a_4 – maintaining the grasp (with force adjustment - the fingers increase/ decrease the squeeze depending on object deformation and slip),
- a_5 – releasing the grasp (the fingers move with a velocity dependent on the knowledge of the object behaviour, e.g. small velocity for an object with an unstable balance),
- a_6 – transition to the rest position (the fingers move with a fixed velocity toward the rest position).

Let us consider the grasping of following objects: a pen and a credit card (standing in a container), a computer mouse and a cell phone (laying on the table), and a kettle and a tube (standing on the table).

For the purpose of simplifying our considerations, the constant time of 256 ms for each action was adopted. This means that a movement of a given type, e.g. closing a spherical grasp (5), is represented by a sequence of the same elementary actions, e.g.: $\langle a_{k+1}^{(5)}, a_{k+2}^{(5)}, \dots, a_{k+n}^{(5)} \rangle$, with a variable number of elements proportional to movement duration.

EMG signals registered in a multi-point system [11] on a forearm of a healthy man were used for the recognition of elementary actions. The measurements were taken by means of 6 electrodes at the frequency of 10^3 samples/s. The *rms* values of signals in 256-sample windows (1 feature/channel) and selected harmonics of an averaged frequency spectrum in these windows (8 features/channel) were considered as potential features. This gives a total of 54 features. Finally the

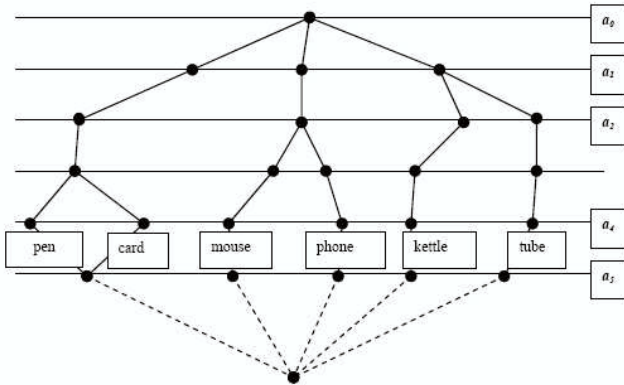


Fig. 2. Exemplary decision tree for patient intent sequence recognition a_0 - rest stage; a_1 - grasp preparation (3 different types of a finger gape); a_2 - grasp closing (different for the kettle and the tube) a_3 - grasping (different for the mouse and the phone - grasp and turn) a_4 - maintaining the grasp (modification of grasp for the pen); a_5 - releasing the grasp (equal for the pen and the card and different for the remaining items); a_6 - transition to the rest position (different for all items)

Table 1. Frequency of correct classification versus the number of learning sequences (NLS) in per cent

NLS	30	40	50	60	70	80	90	100
Result	63.3	66.4	71.3	73.9	77.7	78.6	80.2	84.7

rms values of EMG signals, coming from 3 electrodes were accepted as a feature vector x .

The electrodes were respectively located above the following muscles:

- the wrist extensor (*extensor carpi radialis brevis*);
- wrist flexor (*flexor carpi ulnaris*);
- thumb extensor (*extensor pollicis brevis*).

The basis for determination of classifying functions of recognition algorithm (12) was learning sequences (2) containing a set of pairs: segment of EMG signal/the class of elementary action. Such a set was experimentally determined by means of synchronous registering of movement of fingers and EMG signal. In order to collect learning set a measuring system was used, which concept can be found in [11]. The algorithm was constructed on the basis of the collected learning sequences (1) of the length 7 elementary actions. The tests were conducted on 100 subsequent sequences. The outcome is shown in Table 1. It includes the frequency of correct decisions for the investigated algorithm depending on the number of training sets.

5 Final Remarks

Presented in this paper concept of hand bioprosthesis control is of preliminary nature. On the basis of set of elementary actions for exemplary six stage control procedure, a new recognition system has been proposed. Its idea consists in combination of fuzzy sets and Dempster-Shafer theory into common sequential decision algorithm. Experimental results demonstrate effectiveness of the proposed algorithm in contrast with other methods [10] in the control of bioprosthesis of hand.

Acknowledgement. This work was financed from the Polish Ministry of Science and Higher Education resources in 2007-2010 years as a research project No N518 019 32/1421.

References

1. De Luca, C.J.: The Use of Surface Electromyography in Biomechanics, *Journal of Applied Biomechanics*. *Journal of Applied Biomechanics* 13(2), 128–135 (1997)
2. Eriksson, E., Sebelius, F., Balkenius, C.: Neural control of a virtual prosthesis. In: Vorbrüggen, J.C., von Seelen, W., Sendhoff, B. (eds.) ICANN 1996. LNCS, vol. 1112. Springer, Heidelberg (1996)

3. Kurzynski, M.: Benchmark of Approaches to Sequential Diagnosis. In: Lisboa, P., Ifeachor, J., Szczepaniak, P.S. (eds.) *Perspectives in Neural Computing*, pp. 129–140. Springer, Heidelberg (1998)
4. Kurzyński, M., Wolczowski, A.: Recognition of patient intention via analysis of EMG signal. In: *Proc. 20th Conf. on Biocybernetics and Biomedical Engineering*, Wrocław, 126-130 (2007) (in Polish)
5. Kurzyński, M., Wolczowski, A.: Control of Dexterous Hand via Recognition of EMG Signal Using Combination of Decision-Tree and Sequential Classifier. In: Kurzynski, M., Wozniak, M. (eds.) *Computer Recognition Systems*, vol. 2, pp. 687–694. Springer, Heidelberg (2007)
6. Laschi, C., Dario, P., et al.: Grasping and Manipulation in Humanoid Robotics. In: *First IEEE-RAS Workshop on Humanoids - Humanoids 2000*, Boston (2000)
7. Nishikawa, D.: Studies on electromyogram to motion classifier, Graduate School of Engineering, Hokkaido University, Sapporo, Japa, 2001 (PhD Dissertation) (2001)
8. Wolczowski, A.: Smart Hand: The Concept of Sensor based Control. In: *Proc. of 7th IEEE Int. Symposium on Methods and Models in Automation and Robotics*, Miedzyzdroje (2001)
9. Wolczowski, A., Krysztoforski, K.: Control-measurement circuit of myoelectric prosthesis hand. *Proc. of 13th Conf. of the European Society of Biomechanics, Acta of Bioengineering and Biomechanics 2002* 4 (supp. 1), 576–577 (2002)
10. Wolczowski, A., Kurzynski, M.: Control of Artificial Hand via Recognition of EMG Signals. In: Barreiro, J.M., Martín-Sánchez, F., Maojo, V., Sanz, F. (eds.) *ISBMDA 2004*. LNCS, vol. 3337, pp. 356–367. Springer, Heidelberg (2004)
11. Wolczowski, A., Myśliński, S.: Identifying the relation between finger motion and EMG signals for bioprosthesis control. In: *Proc. of 12th IEEE Int. Conf. on Methods and Models in Automation and Robotics*, Miedzyzdroje (2006)
12. Topolski, M.: Computer sequential classification algorithms combining fuzzy logic and Dempster-Shafer theory. Report PRE 1/07. Wrocław University of Technology (PhD dissertation) (in Polish) (2007)
13. Wierzchon, S.T.: Methods of processing of uncertain information using Dempster-Shafer theory. Institute of Computer Science Polish Academy of Sciences. Warsaw. Poland (in Polish) (1996)

Matching Knowledge and Evidence in a Model of Medical Diagnosis

Ewa Straszecka

Institute of Electronics, Silesian University of Technology
16 Akademicka St, 44-100 Gliwice, Poland
ewa.straszecka@polsl.pl

Summary. The Dempster-Shafer theory extended for fuzzy focal elements can be used to build a flexible model of medical diagnosis. Yet, quality of an inferred diagnosis depends on precision of matching knowledge with evidence (patient's findings). The paper provides definitions of matching precision and suggests methods of the most adequate use of available information about symptoms. The methods are illustrated by an example and tests of an Internet database.

1 Introduction

Various symptoms are analyzed during medical diagnosis. For instance, laboratory tests are numerical variables and they provide crisp input information. On the other hand, linguistic information is used when symptoms found from an interview or primary examination are considered. All symptoms need to be estimated during diagnosis. Depending on a number and an exacerbation of symptoms, certainty of diagnosis should be found. The Dempster-Shafer theory [2, 3, 5] makes it possible to determine belief and plausibility of the diagnosis. An extension of the theory for fuzzy focal elements [7] allow for representation of symptoms by fuzzy sets. Yet, the extension involves the problem of estimation of an accuracy of matching diagnostic knowledge with observed symptoms, i.e. evidence. The accuracy determines credibility of the diagnosis. Hence, it is necessary to use a threshold for an estimation of matching precision. The present paper defines precision of matching and shows methods of the threshold determination. The methods are illustrated by an example and tests performed for Internet data of thyroid gland diseases.

2 The Dempster-Shafer theory

The Dempster-Shafer theory (DST) has been estimated as particularly useful in modeling medical diagnosis [1, 3]. In the theory the basic probability assignment (bpa), denoted as m , is defined for the set A of focal elements a , in the following way [2, 4, 5]:

$$m(f) = 0, \sum_{a \in A} m(a) = 1. \tag{1}$$

where f stands for the false focal element. In a model of medical inference the focal element a describes a symptom or a collection of several symptoms. In the first case it will be called the single while in the second - the complex focal element. On the basis of the bpa the belief and plausibility measures [4, 5] can be determined:

$$Bel(c) = \sum_{(a \Rightarrow c)=t} m(a), \tag{2}$$

$$Pl(c) = \sum_{(a \Rightarrow c) \neq f} m(a), \tag{3}$$

In medical field c can be regarded as a diagnosis. Thus, (2) and (3) are suitable for the representation of belief and plausibility of a diagnostic hypothesis. According to this interpretation, the $[Bel(c), Pl(c)]$ interval corresponds to the certainty of the diagnosis. Attention is focused on belief, i.e. the smaller measure, as conclusions in medicine are drawn very cautiously. When several hypotheses have to be analyzed, their belief measures can be compared. The hypothesis with the greatest Bel value is the final conclusion. If the greatest value occurs with several hypotheses, the final conclusion cannot be determined.

3 Matching Symptoms and Observations

If a symptom is crisp then evidence, i.e. an observation, either entirely matches the focal element or excludes it from inference. On the contrary, when the symptom or the evidence is fuzzy, a partial accuracy of matching is possible. The membership function is a generalization both of the characteristic function and the singleton. Thus, generally, the focal element that regards the x_i variable can be represented by the $\mu(x_i)$ membership function, while the evidence, i.e. observed value of the $i - th$ variable, corresponds to the $\mu^*(x_i)$. Then, the partial accuracy of matching can be defined as the precision level:

$$\eta_l(a_i) = \max_{x_i} \left(\mu_i^{(j,l)}(x_i) \wedge \mu_i^*(x_i) \right), \tag{4}$$

where $\mu_i^{(j,l)}$ concerns the $i - th$ variable, $j - th$ focal element and $l - th$ diagnosis. The $j - th$ complex focal element refers to n^j variables instead of the single variable x_i in (4). In this way a collection of $\eta_l(a_i)$ $i = 1, \dots, n^j$ values is obtained for individual symptoms included in the complex element. The model of diagnosis should base on the most reliable patient cases. Hence, the precision level should be:

$$\eta_l(a_j) = \min_i \eta_l(a_i), a_j = \{a_i\}, i = 1, \dots, n^j. \tag{5}$$

Then, the bpa is defined in the following way:

$$m_l(f) = 0, \quad \sum_{\eta_l(a_j) \geq \eta_{bpa}} m_l(a_j) = 1, \tag{6}$$

where f denotes the focal element: 'lack of symptoms of the $l - th$ diagnosis' and the η_{bpa} is a threshold that determines the 'quality' of knowledge. If the bpa is a normalized frequency of occurrence of symptoms in training data, low η_{bpa} implies knowledge based even on patients cases of doubtful symptoms. If the η_{bpa} is high, than the most clear cases compose knowledge. Yet, it is pointless to set $\eta_{bpa} = 1$, as in this case we step back to the classical definition of focal elements. The (4), (5) definitions of the precision level are also suitable for the belief measure calculation:

$$Bel(D_l) = \sum_{\eta_l(a_j) \geq \eta_T} m_l(a_j). \tag{7}$$

where D_l denotes the $l - th$ diagnosis. By the analogy to (6), the η_T decides how sure evidence takes part is the inference. Yet, when the plausibility measure is regarded, all fuzzy focal elements that match (even partly) evidence should be considered. To this end, let us define the imprecision level:

$$\theta_l(a_j) = \max_i \eta_l(a_i), \quad a_j = \{a_i\}, \quad i = 1, \dots, n^j. \tag{8}$$

Then, the plausibility measure for fuzzy focal elements is:

$$Pl(D_l) = \sum_{\theta_l(a_j) \geq \eta_T} m_l(a_j). \tag{9}$$

The $[Bel(D_l), Pl(D_l)]$ interval is also important in inference, as it may illustrate a difference between what we know for sure and what might be true, when some symptoms would have been confirmed. Medical databases are usually very deficient because of costs (not only economical) of examinations. Thus, the length of the interval is a measure of diagnosis certainty. By means of the introduced definitions, all symptoms of different nature, single or complex may be represented in the model of the diagnosis and can take part in estimation of diagnostic hypotheses. Let us present the model on a simple example.

3.1 Example

Let us assume that we define focal elements for the disease (c) diagnosis in the following way:

- $a_1 \equiv$ "symptom X is present",
- $a_2 \equiv$ "test Y result is high",
- $a_3 \equiv$ "symptom Z is low"
- $a_4 \equiv$ "test Y result is high and symptom Z is normal"
- $f \equiv$ "none of the symptoms is present".

In the same time, the focal elements for the 'health' (h) diagnosis are:

- $b_1 \equiv$ "symptom X is absent",
- $b_2 \equiv$ "test Y result is normal",
- $b_3 \equiv$ "symptom Z is normal"
- $b_4 \equiv$ "test Y result is normal and symptom Z is less than normal"
- $f \equiv$ "none of the symptoms is present",

It is necessary to determine adequate membership functions and the bpas for the both diagnoses to complete knowledge. Let us assume membership functions that are presented in fig.1. The functions for the h diagnosis are denoted by dashed lines, while for the c diagnosis – by dotted lines. Some membership functions of the latter diagnosis are covered by patient observations (solid lines). The bpas can be:

$$m_c(a_1) = 0.3, m_c(a_2) = 0.3, m_c(a_3) = 0.3, m_c(a_4) = 0.1;$$

$$m_h(b_1) = 0.25, m_h(b_2) = 0.3, m_h(b_3) = 0.25, m_h(b_4) = 0.2.$$

The bpa values are defined individually for each diagnosis, so that (6) holds true. Let us now consult a patient. Observations for the patient are: 'symptom X is present', 'test Y result is 2.25' and 'symptom Z is normal' (see solid lines in fig.1). The observations are matched with focal elements both for c and h diagnosis and precision as well as imprecision levels are found. A difference between the levels is significant only for the a_4, b_4 elements. Now the η_T should be chosen and compared with $\eta_c(a_i), \theta_c(a_i)$ as well as with $\eta_h(b_i), \theta_h(b_i), i = 1, \dots, 4$. Then

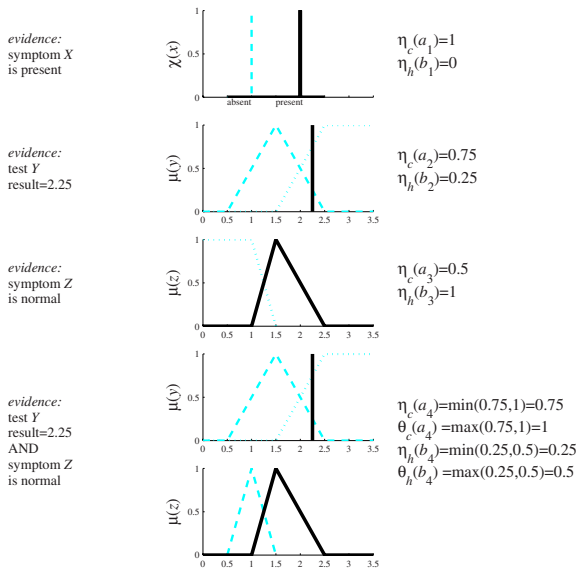


Fig. 1. Matching knowledge with evidence

belief (7) and plausibility (9) values are calculated. The values depend on the η_T threshold:

for $\eta_T = 0.25$

$$Bel(c) = m_c(a_1) + m_c(a_2) + m_c(a_3) + m_c(a_4) = 1, Pl(c) = 1;$$

$$Bel(h) = 0 + m_h(b_2) + m_h(b_3) + m_h(b_4) = 0.75; Pl(h) = 0.75.$$

for $\eta_T = 0.5$

$$Bel(c) = m_c(a_1) + m_c(a_2) + m_c(a_3) + m_c(a_4) = 1, Pl(c) = 1;$$

$$Bel(h) = 0 + 0 + m_h(b_3) + 0 = 0.25,$$

$$Pl(h) = 0 + 0 + m_h(b_3) + m_h(b_4) = 0.45.$$

Thus the problem of the η_T threshold choice is crucial.

3.2 Choice of Inference Threshold

The example shows that the higher is the η_T , the smaller are the values of belief and plausibility. Moreover, the interval $[Bel(h), Pl(h)]$ is also changed. The final diagnosis can be taken after a comparison of the $[Bel(D_l), Pl(D_l)]$ $l = 1, \dots, N$, where N is the number of hypotheses. Still, such a comparison is hardly applicable, as users of diagnosis support systems are rarely familiar with evaluation of intervals. A better idea is to contrast beliefs for different hypotheses. The smaller measure ensures cautious judgment of symptoms and a comparison of its values is easy. If the maximal value of the belief measure is single, then the final conclusion is the diagnosis for which the maximal value occurs. In other case the final conclusion cannot be stated. Anyway, the latter criterion is also influenced by the η_T threshold of inference. In the example $Bel(c) > Bel(h)$ regardless the threshold. Nevertheless, it may happen that for one threshold belief values differ, while for the other, they are equal. Hence, it must be decided which threshold is the most adequate. It is possible to test a number of η_{bpa} and η_T values for training data to find out the most suitable couple of thresholds. Still, this method requires time-consuming calculations. Another approach may be suggested. When knowledge is gathered, rather clear examples of diagnoses are represented. That is why, η_{bpa} should be quite high (but smaller than 1, as it has already been noticed). During inference all possible evidence should be used. Thus, the η_T should be as high as possible, but ensure that the symptom with the least precision level (4) is still considered. The value of the threshold can be found as the highest η_T value, for which $Pl(D_l)$ still remains the maximal possible. In this way η_T becomes also an indicator of quality of the final diagnosis. The both approaches has been tested for an Internet database.

4 Tests

During tests the Internet database <ftp.ics.uci.edu/pub/machine-learning-databases/thyr> oid-disease, files new-thyr have been used. The data concern thyroid

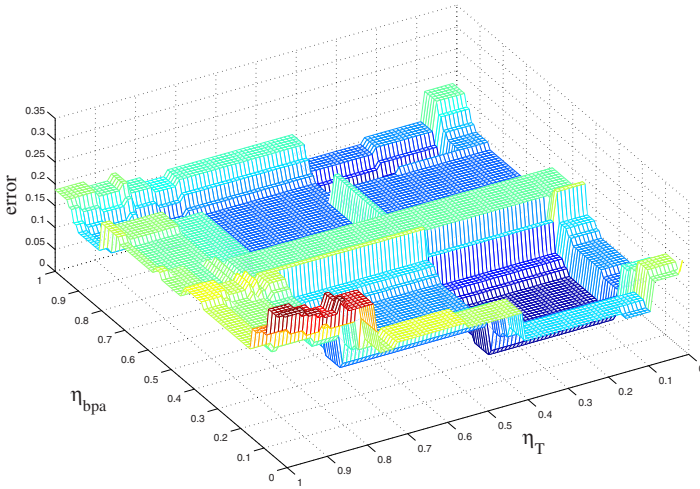


Fig. 2. Global errors for η_{bpa} and η_T couples

gland diseases. Three diagnostic categories (D_1, D_2, D_3) have been considered. The database includes values of 5 laboratory test results for which a diagnosis is stated. The data have been divided for learning/test sets with the following number of cases: 75/75 for D_1 , 15/20 for D_2 , and 15/15 for D_3 . The detailed description of methods used during building of the diagnosis model can be found in references [6, 7]. Here, only the problem of thresholds and $[Bel(D_l), Pl(D_l)]$ intervals is discussed. For the database 5 single focal elements, 3 two-variable and 1 three-variable complex focal elements have been defined. After membership function construction [6], the thresholds have been tested. The η_{bpa} and η_T has been changed in the $[0, 1]$ interval with 0.01 step. Obviously, the η_{bpa} has been used to find the bpa (during training), while the η_T has been used to infer diagnoses for test cases. For each couple of the thresholds the global error has been found as the quotient of number of wrong classified cases and the overall number of cases, for the three diagnoses. Results are presented in the fig.2.

Minimal errors areas have been detected for $\eta_{bpa} \in [0.1, 0.4] \cup [0.8, 0.9]$ and $\eta_T \in [0.2, 0.4]$, yet the thresholds usually have different values ($\eta_{bpa} \neq \eta_T$). The minimum error has been 2.67%. Thus, detection of the appropriate couple of the threshold is not easy. On the contrary, when the η_T threshold has been chosen by the criterion of maximum plausibility, it has been possible to find immediately the right classification for almost each test case (which has resulted in the minimal error of the 2.67%). In that case calculations are shorter, as η_{bpa} is constant and instead of scanning the whole $[0, 1]$ interval for η_T , only several threshold values are tried. It has been also observed that the length of the $[Bel(D_l), Pl(D_l)]$ interval is greater for the non-optimal choice of the thresholds. The fig.3 presents the values of $Cert(D_l) = Pl(D_l) - Bel(D_l)$ $l = 1, 2, 3$, as well as mean values of $Cert$, obtained for test data of the D_1 diagnosis. For the η_{bpa}

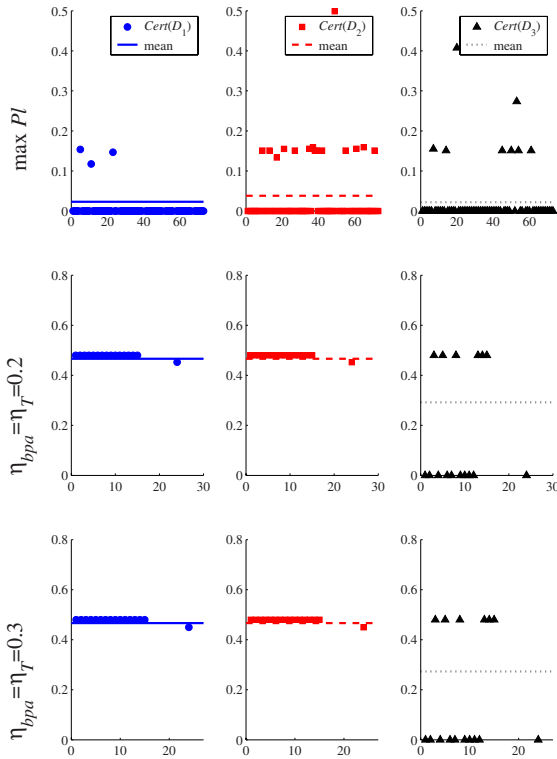


Fig. 3. Diagnosis certainty $Cert(D_i) = Pl(D_i) - Bel(D_i)$ for different thresholds. Mean values of $Cert(D_i)$ are represented by horizontal lines in the subplots. Numbers of correctly classified cases are denoted on the horizontal axes.

and η_T both equal 0.2 or 0.3 certainty values for correctly diagnosed cases are great. The most appropriate thresholds (found for maximal plausibility) result not only in correct diagnoses for almost all cases, but also in smaller $Cert(D_i)$ values. It means that all possible information has been used to determine a diagnosis. Hence, if all available information about symptoms has been used, the diagnosis has been improved.

5 Conclusions

Years of research have proved that the Dempster-Shafer theory is a remarkable alternative to the classical probability theory, particularly in case of medical diagnosis models. Nevertheless, a model of diagnosis that is created in the theory should represent all kind of symptoms: crisp and fuzzy. This requires an extension of the theory for fuzzy focal elements. However, the extension involves necessity of threshold choice, both for basic probability assignment calculation and for inference. It is suggested to choose the former threshold as rather high, to

ensure the right quality of knowledge used for building diagnosis model. Yet, the threshold should be smaller than 1, otherwise symptoms will be interpreted as crisp. The inference threshold can be found when belief and plausibility measures are simultaneously considered. The belief measure is necessary to find possibility of diagnosis on the basis of clearly known symptoms. The plausibility measure indicates credibility of the diagnosis if some more information would be provided. Medical diagnosis is always based on uncertain information, as only necessary examinations are completed for a patient. Unknown symptoms are usually treated in the same way as absent findings, which disturbs the diagnosis. Thus, it is proposed to choose the inference threshold as the highest among the thresholds that are low enough to keep maximum plausibility. This seems to be a good idea to estimate entire possible evidence. The value of the threshold together with the length of the $[Bel, Pl]$ interval are useful indicators of diagnosis credibility. The presented solution is numerically simple and can be easily implemented in a diagnosis support system.

References

1. Beynon, M., Curry, B., Morgan, P.: The Dempster-Sahfer theory of evidence: an alternative approach to multicriteria decision modelling. *Omega* 28, 37–50 (2000)
2. Dempster, A.P.: A generalisation of Bayesian inference. *J. Royal Stat. Soc.*, 205–247 (1968)
3. Gordon, J., Shortliffe, E.H.: The Dempster-Shafer theory of evidence. In: Buchanan, B.G., Shortliffe, E.H. (eds.) *Rule-Based Expert Systems*, pp. 272–292. Addison-Wesley, Reading (1984)
4. Kacprzyk, J., Fedrizzi, M.: *Advances in Dempster-Shafer Theory of Evidence*. Wiley, New York (1994)
5. Shafer, G.: *A mathematical theory of evidence*. Princeton University Press, Princeton (1976)
6. Straszeka, E.: An interpretation of focal elements as fuzzy sets. *Int. J. of Intelligent Systems* 18, 821–835 (2003)
7. Straszeka, E.: Combining uncertainty and imprecision in models of medical diagnosis. *Information Sciences* 176, 3026–3059 (2006)

Nonparametric Regression for Analyzing Correlation between Medical Parameters

Malgorzata Charytanowicz and Piotr Kulczycki

Polish Academy of Sciences, Systems Research Institute, ul. Newelska 6,
PL-01-447 Warsaw
malgorzata.charytanowicz@ibspan.waw.pl

Summary. Chronic renal failure is associated with major biochemical and hematological derangements. These changes are often represented as linear functions of creatinine. The aim of the study is to analyze the correlation of hematologic parameters with creatinine. The sample population involved patients with renal insufficiency observed in the Stefan Kardynal Wyszyński Regional Specialists' Hospital in Lublin (Poland). The method presented here is based on the theory of statistical kernel estimators, which frees it of assumptions in regard to the form of regression function.

1 Introduction

Chronic renal failure (CRF) is characterized by a slow, progressive decline in glomerular filtration rate which increasingly effects other kidney functions. Progression of renal disease is associated with development of anemia. A negative correlation between creatinine and hematologic parameters (hemoglobin, hematocrit, red blood cells) exists. Inhibitors of erythropoiesis have been suggested to be important in the pathogenesis of anemia. Inadequate erythropoiesis occurs because the quantity of endogenous erythropoietin produced by the peritubular fibroblasts in the kidney, is insufficient in relation to the degree of anemia. Patients with CRF are unable to increase erythropoiesis sufficiently to compensate blood loss [8, 9]. Correction of anemia becomes possible due to the availability of recombinant human erythropoietin (rHuEPO) [11].

The goal of this paper is to provide a method allowing the analysis the relationship of the hematological parameters, performed separately, and creatinine. Creatinine was chosen as a marker of renal insufficiency. The data was derived from patients with renal insufficiency. The proposed method allows determine the stage of renal failure indicated the beginning of anemia. The mathematical apparatus relies on the theory of statistical kernel estimators [5, 6, 7, 10], which frees the method from the types of regression functions. The preliminary version of this paper was presented as [1].

2 Materials and Methods

The sample population involved patients with renal insufficiency who had not received rHuEPO, observed in the Stefan Kardynal Wyszyński Regional Specialists' Hospital in Lublin (Poland), for periods up to five years (1994-1998). During this time, determinations of creatinine, as well as other biochemical and hematological parameters, were done routinely. The study is composed of 1915 determinations in the creatinine range from 1.1 to 5.0 mg/dL. The normal range for creatinine values equals 0.7-1.4 mg/dL. The number of observations with creatinine exceeded normal values equals 908. The mean, standard deviation and median values of standard blood parameters, including hemoglobin, hematocrit and red cells for this group are shown in Table 1.

The mean \pm S.D. of hemoglobin concentrations were 13.28 ± 2.63 g/dL with the median of 13.50 g/dL for males, and 11.91 ± 1.93 g/dL with the median of 11.81 g/dL for females. The mean \pm S.D. of hematocrit were 38.37 ± 7.57 % with the median of 38.70 % for males, and 34.43 ± 5.43 % with the median of 34.25 % for females. The mean \pm S.D. of red blood cells count were 4.33 ± 0.87 [$\times 10^{12}$ /L] with the median of 4.37 [$\times 10^{12}$ /L] for males, and 3.84 ± 0.61 [$\times 10^{12}$ /L] with the median of 3.84 [$\times 10^{12}$ /L] for females. In fact, all of the following parameters are lower than the normal ranges given in Table 2.

In treated patients, there was a significantly negative correlation between the hemoglobin concentration and creatinine ($r = -0.45, p < 0.0001$ for males, $r = -0.49, p < 0.0001$ for females). The correlation was significantly negative between hematocrit and creatinine ($r = -0.46, p < 0.0001$ for males, $r = -0.49, p < 0.0001$ for females) and also between the red blood cells count and creatinine ($r = -0.49, p < 0.0001$ for males, $r = -0.50, p < 0.0001$ for females). The rate of hemoglobin, hematocrit and red blood cells decreases when

Table 1. Mean (\pm S.D.) and median values of selected hematologic parameters for creatinine range ≤ 5.0 mg/dL

Parameters	Male ($n = 534$)		Female ($n = 374$)	
	Mean \pm S.D.	Median	Mean \pm S.D.	Median
Hemoglobin [g/dL]	13.28 ± 2.63	13.50	11.91 ± 1.93	11.81
Hematocrit [%]	38.37 ± 7.57	38.70	34.43 ± 5.43	34.25
Red Blood Cells [$\times 10^{12}$ /L]	4.33 ± 0.87	4.37	3.84 ± 0.61	3.84

Table 2. Normal range of selected hematologic parameters

Parameter	Male	Female
Hemoglobin [g/dL]	14.00–18.00	12.00–16.00
Hematocrit [%]	42.00–52.00	37.00–47.00
Red Blood Cells [$\times 10^{12}$ /L]	4.70–6.10	4.20–5.40

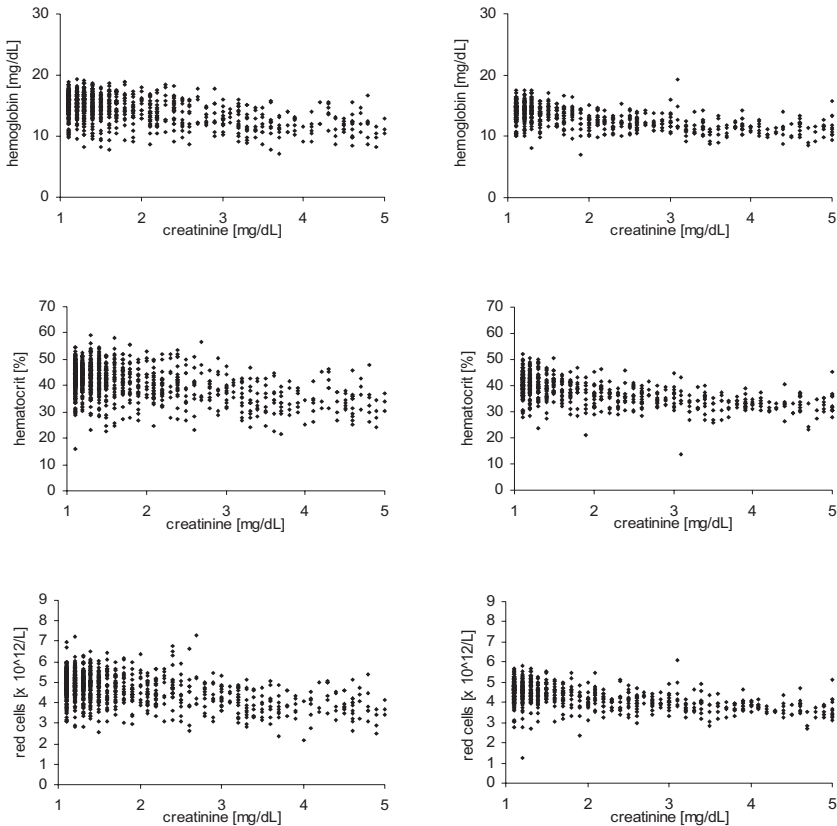


Fig. 1. Scatterplot of creatinine and selected hematologic parameters for males and females

creatinine increases. Scatter diagrams given on Figure 1 shows selected hematologic parameters as functions of creatinine, separately for males (first column) and females (second column).

During early renal failure (creatinine < 5 mg/dL) hematological parameters have often been represented as linearly related to creatinine [3]. The kernel-based nonparametric regression estimators give a much more flexible family of curves to choose from.

3 Nonparametric Regression

3.1 Kernel Regression

Classical parametric methods of determining an appropriate functional relationship between the two variables imposed arbitrary assumptions concerning the

functional form of the regression function. The choice of parametric model depends very much on the situation. If a chosen parametric family is not of appropriate form, then there is a danger of reaching incorrect conclusions in the regression analysis. This also makes it difficult to take into account the whole accessible information. The rigidity of this regression can be overcome by removing the restriction that the model is parametric. This approach leads to nonparametric regression that let the data decide which function fits them best. In this study, a class of kernel-type regression estimators called local polynomial kernel estimators is presented.

Let therefore, n elements $(x_i, y_i) \in \mathbb{R} \times \mathbb{R}$, $i = 1, 2, \dots, n$ be given, where values x_i may designate some non-random numbers or realizations of the one-dimensional random variable X , whereas y_i designate realizations of the one-dimensional random variable Y . Assuming the existence of the function $f : \mathbb{R} \rightarrow \mathbb{R}$ having a continuous first derivative that:

$$y_i = f(x_i) + \varepsilon_i, \tag{1}$$

where ε_i are independent random variables with zero mean and unit, finite variance. Let then $p \in \mathbb{N}$ be the degree of the polynomial being fit. The kernel regression estimator $\hat{f} : \mathbb{R} \rightarrow \mathbb{R}$, obtained by using weighted least squares with kernel weights, can be given by the formula:

$$\hat{f}(x) = e_1^T (X^T W X)^{-1} X^T W y, \tag{2}$$

where

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{3}$$

means the vector of responses,

$$X = \begin{bmatrix} 1 & x_1 - x & (x_1 - x)^2 & \dots & (x_1 - x)^p \\ 1 & x_2 - x & (x_2 - x)^2 & \dots & (x_2 - x)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n - x & (x_n - x)^2 & \dots & (x_n - x)^p \end{bmatrix} \tag{4}$$

is an $n \times (p + 1)$ design matrix,

$$W = \text{diag} \left(\frac{1}{h} K \left(\frac{x_1 - x}{h} \right), \frac{1}{h} K \left(\frac{x_2 - x}{h} \right), \dots, \frac{1}{h} K \left(\frac{x_n - x}{h} \right) \right) \tag{5}$$

denotes an $n \times n$ diagonal matrix of kernel weights,

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{6}$$

is the $(p + 1) \times 1$ vector having 1 in the first entry and zero elsewhere. The coefficient $h > 0$ is called a bandwidth, while the measurable function $K : \mathbb{R} \rightarrow [0, \infty)$ of unit integral, symmetrical with respect to zero and having a weak global maximum in this place, takes the name of the kernel.

The choice of the kernel form has no practical meaning and thanks to this, it is possible to take into account the primarily properties of the estimator obtained. Most often the standard normal kernel expressed by a convenient analytical formula:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{7}$$

is used.

The practical implementation of the kernel regression estimators requires a good choice of bandwidth. If h is too small, a spiky rough kernel estimate is obtained, and if h is too large, it results a flat kernel estimate. A frequently used bandwidth selection technique is the cross-validation method [2, 4], which chooses h to minimize

$$\sum_{i=1}^n \left(y_i - \hat{f}_{-i}(x_i) \right)^2 \tag{8}$$

where $\hat{f}_{-i}(\cdot)$ denotes the regression estimator (2), without using the i th observation (x_i, y_i) .

An important problem is the choice of the parameter p . For sufficiently smooth regression functions, the asymptotic performance of \hat{f} improves for higher values of p . However, for higher p , the variance of the estimator becomes larger and in practice, a very large sample may be required. On the other hand, the even degree polynomial kernel estimator has a more complicated bias expression which does not lend itself to simple interpretation. These facts suggests the use of either $p = 1$ or $p = 3$. Moreover, for $p = 1$, the convenient explicit formulae exists:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n \frac{(\hat{s}_2(x) - \hat{s}_1(x)(x_i - x)) y_i K\left(\frac{x-x_i}{h}\right)}{\hat{s}_2(x)\hat{s}_0(x) - (\hat{s}_1(x))^2} \tag{9}$$

where

$$\hat{s}_r(x) = \frac{1}{nh} \sum_{i=1}^n (x_i - x)^r K\left(\frac{x_i - x}{h}\right) \quad r = 0, 1, 2. \tag{10}$$

Therefore - except in more advanced statistical applications - $p = 1$ is preferred. The tasks concerning the choice of the kernel form, the bandwidth, as well as additional procedures improving the quality of the estimator obtained, are found in [5, 6, 10]. The utility of local linear kernel estimators has been investigated in the context of some typical data derived from patients with renal insufficiency.

3.2 Results

In this study, the method of nonparametric regression based on a weighted local linear regression was used to analyze the relationship between creatinine and selected hematological parameters (hematocrit, hemoglobin and red cells). For ease

of computation, the standard normal kernel (7) was used. The bandwidth was determined using the cross-validation method (8). The results were compared with results obtained for two parametric models:

- linear regression model [3]

$$y = b_0 + b_1x \tag{11}$$

- logarithmic regression model

$$y = b_0 + b_1 \log x \tag{12}$$

having the best fit to the data in the family of nonlinear functions.

The Mean Squared Error *MSE*

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{13}$$

precisely, the Root of the Mean Squared Error *RMSE* was reported in comparison with the weighted local linear regression (9). Table 3 shows estimation errors *RMSE* obtained for these three considered regression models describing the relationship of hemoglobin, hematocrit and red cells to creatinine, separately for males and females.

Table 3. Comparing of the estimation errors for selected regression functions describing the relationship of hematological parameters (hemoglobin, hematocrit, red blood cells) to creatinine for males ($n = 1239$) and females ($n = 676$)

Male		Female	
Regression Equation	<i>RMSE</i>	Regression Equation	<i>RMSE</i>
Hemoglobin <i>HB</i> [g/dL]			
<i>HB</i> = 16.75 – 1.20· <i>creatinine</i>	1.77	<i>HB</i> = 15.08 – 0.95· <i>creatinine</i>	1.41
<i>HB</i> = 15.84 – 2.63·log(<i>creatinine</i>)	1.77	<i>HB</i> = 14.45 – 2.27·log(<i>creatinine</i>)	1.44
kernel estimator (9)	1.73	kernel estimator (9)	1.35
Hematocrit <i>HT</i> [%]			
<i>HT</i> = 47.38 – 3.14· <i>creatinine</i>	5.23	<i>HT</i> = 43.04 – 2.56· <i>creatinine</i>	4.10
<i>HT</i> = 44.98 – 6.82·log(<i>creatinine</i>)	5.24	<i>HT</i> = 41.34 – 6.06·log(<i>creatinine</i>)	4.05
kernel estimator (9)	5.12	kernel estimator (9)	3.92
Red Blood Ceels <i>RBC</i> [$\times 10^9$ /L]			
<i>RBC</i> = 5334.48 – 347.01· <i>creatinine</i>	6.15	<i>RBC</i> = 4857.05 – 274.40· <i>creatinine</i>	4.89
<i>RBC</i> = 5067.61 – 750.08·log(<i>creatinine</i>)	6.17	<i>RBC</i> = 4671.56 – 641.49·log(<i>creatinine</i>)	4.86
kernel estimator (9)	6.02	kernel estimator (9)	4.69

Table 4. Calculated values of creatinine corresponding to the values less than the lower limit of normal ranges of selected hematologic parameters

Parameter	Creatinine [mg/dL]	
	Male	Female
Hemoglobin [g/dL]	2.11	2.76
Hematocrit [%]	2.02	2.32
Red Blood Cells [$\times 10^{12}/L$]	2.15	2.63

The *RMSE* for both the linear and logarithmic functions are comparable. The mean square error for males is a bit greater than for females. The local linear estimator gives the lowest mean squared errors and the difference is in the range from 2% to 4%.

Further analysis was performed using kernel estimators of regression. The functional form of the relationship of all dependent variables is determined as a function of creatinine. Using obtained forms, the values of creatinine corresponding to the values less than the lower limit of normal ranges of hemoglobin concentrations, hematocrit and red blood cells count were calculated. The results, obtained separately for males and females, are given in Table 4.

The analysis of various hematological parameters has allowed the claim that anemia in chronic renal failure develops when creatinine exceed 2 mg/dL for males and 2.3 mg/dL for females. The dispersion of all these parameters is greater in the male's group. The first step in the correction of renal anemia is confirmation of the diagnosis. This diagnosis should be done before rHuEPO treatment. If renal anemia appears, other causes of anemia such as blood loss, iron deficiency or malnutrition should be sought and corrected. Treatment decisions for patients with renal anemia should be recommended at an early stage, before symptoms appear.

4 Summary

Anemia is a common complication of chronic renal failure (CRF). Over fifty percent of patients with CRF die because of anemia. It has been reported that dialysis have no significant meaning in treatment of anemia in patients with renal disease. In eighty years, much progress has been made in the management of anemia due to the availability of recombinant human erythropoietin (rHuEPO). The clearest benefit of rHuEPO in CRF is a substantial reduction in transfusion dependency, which reduces the need for hospital admission and the risk of viral transmission. Treatment of anemia with rHuEPO has also been shown to improve cognitive function, socialization and quality of life in dialysis patients and in consequence, is required and beneficial. To obtain the maximum benefits of early anemia correction, physical training and diet are necessary.

The aim of this study was to analyze the relationship between hematological parameters and stage of kidney disease. A negative correlation between creatinine

and hematocrit, creatinine and hemoglobin and also between creatinine and red cells was observed. The proposed procedure, based on the methodology of kernel estimators, has allowed the claim that anemia in renal insufficiency develops when creatinine exceed 2 mg/dL. The kernel regression method enabled better use of the available data and more sophisticated analysis.

References

1. Charytanowicz, M., Kulczycki, P. (2007) XV Krajowa Konferencja Naukowa - Biocybernetyka i Inżynieria Biomedyczna, Wrocław, 12-15, CD: 025 (September 2007)
2. Fan, J.: *Journal of the American Statistical Association* 87, 998–1004 (1992)
3. Hakim, R.M., Lazarus, J.M.: *American Journal of Kidney Diseases* XI, 238–247 (1988)
4. Hardle, W., Marron, J.S.: *The Annals of Statistics* 13, 1465–1481 (1985)
5. Kulczycki, P.: *Estymatory jadrowe w analizie systemowej*. WNT, Warsaw (2005)
6. Kulczycki, P., Charytanowicz, M.: *Applied Mathematics and Computer Science* 15, 393–404 (2005)
7. Kulczycki, P., Charytanowicz, M.: *Cybernetics and Systems, Asymmetrical Conditional Bayes Parameter Identification for Control Engineering* (in press) (2008)
8. Orłowski, T.: *Choroby nerek*. PZWL, Warsaw (1992)
9. Popławski, A.: *Przegląd Lekarski* 52, 78–79 (1995)
10. Wand, M.P., Jones, M.C.: *Kernel Smoothing*. Chapman and Hall, London (1994)
11. Winearles, C.G., Oliver, D.O., Pippard, M.J., Reid, C., Downing, M.R., Cotes, P.M.: *Lancet*. 2, 1175–1182 (1986)

Experiments on Linear Combiners

Michał Wozniak

Chair of Systems and Computer Networks, Wrocław University of Technology,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
michal.wozniak@pwr.wroc.pl

Summary. The Multiple Classifier Systems are nowadays one of the most promising directions in pattern recognition. There are many methods of decision making by the group of classifiers. The most popular are methods that have their origin in vote methods, where the decision of the common classifier is a combination of simple classifiers decisions. On the other hand there exists a trend of combined classifiers, which are making their decisions basing on the discrimination function, this function is a combination of above-mentioned simple classifier functions. This work presents an attempt to estimate the classifier error, which bases on the combined discrimination function. Obtained from this estimation conclusions will serve to formulate project guidelines for this type of decision-making systems. At the end experimental results of combining algorithms are presented.

1 Introduction

The concept of the Multiple Classifier Systems (MCS) is not new and it is known for over 15 years [18]. Some works in this field were published as early as the '60 of the XX century [2], when it was shown that the common decision of independent classifiers is optimal, when chosen weights are inversely proportional to errors made by classifiers. In many review articles this trend is mentioned as one of the most promising in field of the pattern recognition [10]. In the beginning in literature one could find only the majority vote, but in later works more advanced methods of receiving common answer of the classifier group were proposed. Attempts to estimate classification quality by the classifier committee are one of essential trends. Known in literature conclusions, derived on the analytic way, concern particular case of the majority vote [8], when classifier committee is formed from independent classifiers. Unfortunately this case has only theoretical character and is not useful in practice. On the other hand the weighted vote is taken into consideration [13]. In this work it was pointed that the optimal weight value should be dependent on the error of the simple classifier and on the *prior* probability of the class, on which classifier points. When it comes to combined systems, which base their decisions on the common value of the discrimination function, one has to list the work [15], in which the optimal projective fuser was presented. One also has to mention many other works, that describe analytical properties and experimental results, like [1, 6, 9, 16]. In this work we consider

only fusion of classifiers on the level of their discrimination functions, which in a case of the Bayes algorithm is the *posterior* probability. This work also presents the upper bound estimation of the error made by each classifier in relation to the optimal Bayes classifier. Above-mentioned estimation is used for upper bound estimation of the error made by the classifier committee that for making decision uses the *posterior* probability estimator. This *posterior* probability estimator is a linear combination of values of simple classifier estimators. Our observations are verified by experimental results on computer generated data.

2 Problem Statement

To begin let's present briefly the problem of the recognition in Bayes' decision theory and the fusion method of the classifier's answer. Among different approaches to the uncertainty management in computer-aided recognition systems the statistical decision theory is still an attractive and effective approach. This theory assumes [4] that both the feature vector x and the class number $\omega_j \in \{\omega_1, \omega_2, \dots, \omega_c\}$, which object belongs to, are realizations of the pair of random variables X, J . They are described using the *prior* probability of classes $P(\omega_j)$ and using the class conditional probability density function $p(x | \omega_j)$, for each of the classes. Above characteristics serve for deriving the *posterior* probability, which in particular case (for loss function 0-1) acts as the discrimination function [3] of the optimal Bayes classifier Ψ^*

$$\Psi^*(x) = \omega_i \text{ if } P(\omega_i | x) = \max_{k \in \{1, \dots, c\}} P(\omega_k | x) \tag{1}$$

where

$$P(\omega_i | x) = \frac{P(\omega_i)p(x | \omega_i)}{\sum_{k=1}^c P(\omega_k)p(x | \omega_k)}. \tag{2}$$

In real decision problems above characteristics are unknown and our knowledge of them is hidden in given learning information. Our goal is then the estimation of unknown characteristics and usage of derived estimators for making decision according to the Bayes' rule. Let's assume that we have n classifiers $\Psi^{(1)}, \Psi^{(2)}, \dots, \Psi^{(n)}$. Each of them makes decision basing on the estimation of the *posterior* probability. Let $P^{(l)}(\omega_i | x)$ denotes the *posterior* probability estimator of the class i with given value of x . This estimator is used by the l th classifier $\Psi^{(l)}$. For making common decision by the group of classifiers $\Psi^{(1)}, \Psi^{(2)}, \dots, \Psi^{(n)}$ let's use following classifier $\bar{\Psi}$

$$\bar{\Psi}^*(x) = \omega_i \text{ if } \bar{P}(\omega_i | x) = \max_{k \in \{1, \dots, c\}} \bar{P}(\omega_k | x), \tag{3}$$

where

$$\bar{P}(\omega_i | x) = \sum_{l=1}^n \alpha^{(l)} P^{(l)}(\omega_i | x) \text{ and } \sum_{l=1}^n \alpha^{(l)} = 1. \tag{4}$$

3 Analytical Estimation of Upper Errors

Let's note that according to the Bayes' decision theory each classifier has an error not smaller than the error of the optimal classifier (1) [3]. Let's try to estimate the error made by the classifier (3) in relation to the optimal classifier. For this purpose let's formulate the theorem that estimates the upper bound of the error made by any classifier (let's name it $\Psi^{(l)}(x)$). This classifier for making decision uses the *posterior* probability estimator $P^{(l)}(\omega_i | x)$. Let's also assume that this estimator is derived with the accuracy $\epsilon^{(l)}(x)$, i.e.

$$\forall x \in X \wedge \forall i \in \{1, \dots, c\} | P^{(l)}(\omega_i | x) - P(\omega_i | x) \leq \epsilon^{(l)}(x) \tag{5}$$

Let $P_e^{(l)}$ denotes the $\Psi^{(l)}$ classifier error probability and P_e denotes the Bayes classifier error probability.

Theorem 1

$$P_e^{(l)} - P_e \leq \int_X 2\epsilon^{(l)}(x)p(x)dx, \tag{6}$$

where $p(x)$ denotes unconditional probability density of x

$$p(x) = \sum_{k=1}^c P(\omega_k)p(x | \omega_k). \tag{7}$$

Proof

Let's consider the worse case where classifier makes decision different than the optimal classifier

$$\begin{aligned} \forall x \in X \quad P^{(l)}(\omega_i^{(l)} | x) &= \max_{k \in \{1, \dots, c\}} P^{(l)}(\omega_k | x) \\ \wedge P(\omega_i^* | x) &= \max_{k \in \{1, \dots, c\}} P(\omega_k | x) \wedge \omega_i^{(l)} \neq \omega_i^* \end{aligned} \tag{8}$$

Then the considered classifier, for given value of the feature vector x , makes an error $P_e^{(l)}(x) = 1 - P(\omega_i^{(l)} | x)$. If considered classifier makes decision different than the Bayes classifier, i.e. estimator used by the considered classifier is higher for class other than in case of optimal classifier. Let $\omega_i^{(l)}$ denotes class number pointed by algorithm $\Psi^{(l)}$. On the basis of assumptions we made

$$P(\omega_i^{(l)} | x) - \epsilon^{(l)}(x) \leq P^{(l)}(\omega_i^{(l)} | x) \leq P(\omega_i^{(l)} | x) + \epsilon^{(l)}(x), \tag{9}$$

therefore

$$P_e^{(l)}(x) \leq 1 - P^{(l)}(\omega_i^{(l)} | x) + \epsilon^{(l)}(x). \tag{10}$$

On the other hand for the same value x the Bayes algorithm points at class ω_i^* and following inequality occurs

$$P(\omega_i^* | x) - \epsilon^{(l)}(x) \leq P^{(l)}(\omega_i^* | x) \leq P(\omega_i^* | x) + \epsilon^{(l)}(x), \tag{11}$$

Making use of the fact that the error probability for Bayes classifier is given by $P_e(x) = 1 - P(\omega_i^* | x)$, where ω_i^* is the class number pointed by the Bayes algorithm, we finally have

$$P_e(x) \geq 1 - P^{(l)}(\omega_i^* | x) - \epsilon^{(l)}(x). \tag{12}$$

Equating the inequalities (10) and (12) we derive the upper bound estimation of the classifier error in relation to the Bayes classifier

$$\begin{aligned} P_e^{(l)} - P_e &\leq 1 - P^{(l)}(\omega_i^{(l)} | x) + \epsilon^{(l)}(x) - 1 + P^{(l)}(\omega_i^* | x) + \epsilon^{(l)}(x) = \\ &= 2\epsilon^{(l)}(x) + P^{(l)}(\omega_i^* | x) - P^{(l)}(\omega_i^{(l)} | x). \end{aligned} \tag{13}$$

Let's also note that since classifier $\Psi^{(l)}$ made decision $\omega_i^{(l)}$, then

$$P^{(l)}(\omega_i^* | x) - P^{(l)}(\omega_i^{(l)} | x) < 0, \tag{14}$$

therefore

$$P_e^{(l)} - P_e \leq 2\epsilon^{(l)}(x) + P^{(l)}(\omega_i^{(l)} | x) - P^{(l)}(\omega_i^{(l)} | x) \leq 2\epsilon^{(l)}(x). \tag{15}$$

Let's then derive the upper bound estimation of the classifier's average error

$$P_e^{(l)} - P_e = E_X[p_e^{(l)} - p_e(x)] \tag{16}$$

Making use of (15) we have

$$P_e^{(l)} - P_e = \int_X 2\epsilon^{(l)}(x)p(x)dx. \tag{17}$$

Let's use above theorem for estimating the error of classifier (3).

Theorem 1. *Theorem 2 Let \bar{P}_e denotes the average probability of error of the classifier, which makes a decision according to (4). Upper estimation of the difference between the above-mentioned error and the Bayes classifier error is given by the formula*

$$\bar{P}_e - P_e \leq 2 \int_X \sum_{l=1}^c \alpha^{(l)} \epsilon^{(l)}(x)p(x)dx. \tag{18}$$

Proof. Since each l th classifier for making decision uses the posterior probability estimator with accuracy $\epsilon^{(l)}(x)$, then the common classifier, which uses the weighted estimator derived according to formula (4), will be derived with accuracy $\sum_{l=1}^N \alpha^{(l)} \epsilon^{(l)}(x)$. The upper bound estimation of the error of the above-mentioned common classifier in relation to the optimal classifier, according to the Theorem 1, will be given by

$$\int_X \sum_{l=1}^N \alpha^{(l)} \epsilon^{(l)}(x)p(x)dx. \tag{19}$$

As a result of the above considerations we derived the upper bound estimation of the error made by the common weighted classifier. Its lower bound estimation is determined by the error of the optimal Bayes classifier. Let's note that the upper bound estimation will not be higher than the estimation of the worst classifier of the group and will be better than the estimation of the best classifier of the group, because

$$\min_{l \in \{1, \dots, N\}} \epsilon^{(l)}(x) \leq \sum_{k=1}^N \alpha^{(k)} \epsilon^{(k)}(x) \leq \max_{l \in \{1, \dots, N\}} \epsilon^{(l)}(x). \quad (19)$$

Above conclusions refer only to the upper bound estimation of the error and does not tell us anything about the classifier quality in relation to simple classifiers. The obstacle here is mostly due to not knowing the error estimation function, $\epsilon(x)$ for each *posterior* probability estimator. Let's only note that it is possible to get classifier, for which the upper bound estimation of the error will be smaller than the upper bound estimation of each simple classifier. For that purpose the weights that determine the power, with what each classifier takes part in the common decision, cannot be constant, but these weights have to be functions of the argument x . Unfortunately as we noted earlier the function $\epsilon(x)$ is unknown for wanted classifier. Therefore selection of the weight functions, $\alpha^{(l)}(x)$ for each classifier should be made independently for each decision task. These functions have a heuristic character and certainly should be derived by skilled experts, who basing on their experience and intuition, should provide such functions. The quality of those functions should be verified by the computer experiments. Let's also note that proposed form of weight coefficients, which suggest taking into consideration, for the given value of x , the best (the most accurate) classifier can be found in literature under name classifiers selection [12, 14, 17].

4 Experimental Investigation

Since it is not possible to determine the quality of above-mentioned classifiers in analytical way, it was decided to carry out computer experiment.

4.1 Conditions of Experiment

The aim of the experiment was to compare the errors of the weighted combined classifiers with the quality of simple classifiers with different qualities of recognition and combined ones that in combining rules do not take into account the qualities of simple classifiers [11, 13]. All experiments were carried out in Matlab environment using the PRtools toolbox [5]. Errors of the classifiers were estimated using the 10-cross-validation method. In all experiments two-class recognition task was considered with the prior probability values equal 0.5 for each classes. The conditional density functions were generated according the Normal or Banana distribution. The examples of learning sets are depicted in Fig.1.

To obtain set of classifier with different qualities of recognition we decided to add to the learning set some elements generated with unperturbed probability

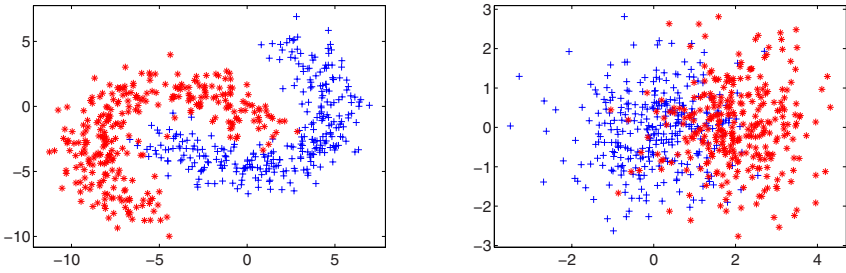


Fig. 1. Examples of banana (left) and Normal (right) distribution

distribution (noise). The quality of classifier which learned on more noised data has worse quality. For estimation of the probability density the $k(n)$ -NN (Nearest Neighbour) estimator [7] and the Parzen estimator were used [3]. For comparison listed below classifiers were chosen:

1. classifiers based on the learning sets generated according to the chosen probability distribution, perturbed 0, 10, 20, 30, 40 % of elements of the uniform distribution – denoted as S1, S2, S3, S4, S5 respectively,
2. majority voting – denoted as MV,
3. weighted voting, where each simple classifier voted with power proportional to $\frac{p}{1-p}$, where p was the classifier’s error TUTAJ denoted as PV,
4. classifier based on weighted function of the *posterior* probability, where the weight of each classifier amounted to 0.2 – denoted as AM,
5. classifier based on weighted function of the *posterior* probability, where the weight of each classifier was proportional to $\frac{p}{1-p}$, where p was the classifier’s error – denoted as WS.

Obtained results are presented in Fig.2-3.

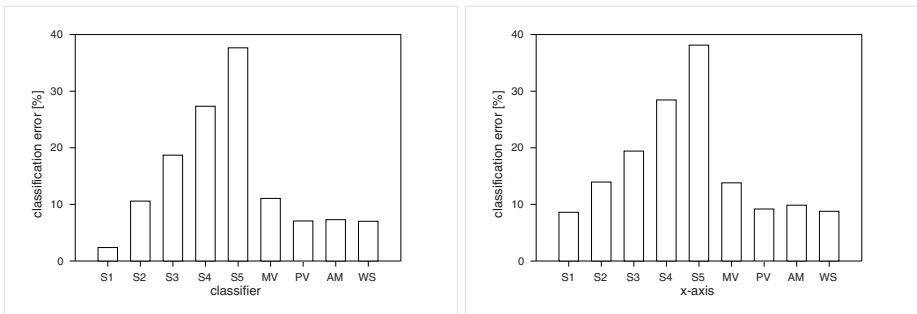


Fig. 2. Frequency of the incorrect classification for banana (left) and Normal (right) distributions and k_n NN probability estimation

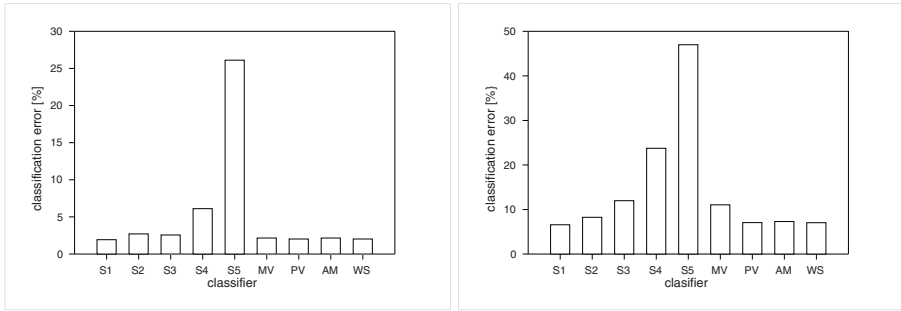


Fig. 3. Frequency of the incorrect classification for banana (left) and Normal (right) distributions and Parzen probability estimation

4.2 Experimental Results Evaluation

Firstly one has to note that we are aware of the fact that the scope of experiments is limited. Therefore making general conclusions basing on them is very risky and in our opinion mentioned below statements should not be generalized at this point, but they should be confirmed by other experiments in much broader scope. In the experiments combined algorithms gave results slightly worse than the best single classifier. It was also observed that classifiers, which for making decision used weighted answer of single classifiers, classified objects definitely better than the rest of combined classifiers.

5 Final Remarks

This study presented the experimental and analytical quality evaluation of the combined classifiers recognition, where these classifiers for making decision used the posterior probability estimators. Obtained results seem to confirm the sense of using combining methods. They also suggest the usage of combining rules, which base on the valuation consistent with the local quality of simple classifiers. Unfortunately, as it was shown in chapter 3, it is not possible to determine their value in the analytical way. One should although hope that in case of application of above mentioned methods for real decision tasks, we have judgment of the expert, who can formulate the heuristic function of weights selection. This function then could be verified and improved by the computer experiment.

References

1. Biggio, B., et al.: Bayesian Analysis of Linear Combiners. In: Haindl, M., Kittler, J., Roli, F. (eds.) MCS 2007. LNCS, vol. 4472, pp. 292–301. Springer, Heidelberg (2007)
2. Chow, C.K.: Statistical independence and threshold functions. IEEE Trans. on Electronic Computers EC(16), 66–68 (1965)

3. Devijver, P.A., Kittler, J.: Pattern Recognition: A Statistical Approach. Prentice Hall, London (1982)
4. Duda, R.O., et al.: Pattern Classification. John Wiley, Chichester (2001)
5. Duin, R.P.W., et al.: PRTools4, A Matlab Toolbox for Pattern Recognition, Delft University of Technology (2004)
6. Fumera, G., Roli, F.: A Theoretical and Experimental Analysis of Linear Combiners for Multiple Classifier Systems. IEEE Trans. on PAMI 27(6), 942–956 (2005)
7. Goldstein, M.: k_n -Nearest Neighbour Classification. IEEE Trans. on IT (September 1972)
8. Hansen, L.K., Salamon, P.: Neural Networks Ensembles. IEEE Trans. on PAMI 12(10), 993–1001 (1990)
9. Hashem, S.: Optimal linear combinations of neural networks. Neural Networks 10(4), 599–614 (1997)
10. Jain, A.K., et al.: Statistical Pattern Recognition: A Review. IEEE Transaction on PAMI 22(1), 4–37 (2000)
11. Kittler, J., Alkoot, F.M.: Sum versus Vote Fusion in Multiple Classifier Systems. IEEE Trans. on PAMI 25(1), 110–115 (2003)
12. Kuncheva, L.I.: Change-glasses approach in pattern recognition. Pattern Recognition Letters 14, 619–623 (1993)
13. Kuncheva, L.I., et al.: Limits on the Majority Vote Accuracy in Classifier Fusion. Pattern Analysis and Applications 6, 22–31 (2003)
14. Kuncheva, L.I.: Combining pattern classifiers: Methods and algorithms. Wiley-Interscience, New Jersey (2004)
15. Rao, N.S.V.: A Generic Sensor Fusion Problem: Classification and Function Estimation. In: Roli, F., Kittler, J., Windeatt, T. (eds.) MCS 2004. LNCS, vol. 3077, pp. 16–30. Springer, Heidelberg (2004)
16. Tumer, K., Ghosh, J.: Linear and Order Statistics Combiners for Pattern Classification. In: Sharkley, A.J.C. (ed.) Combining Artificial Neural Networks, pp. 127–155. Springer, Heidelberg (1999)
17. Woods, K., et al.: Combination of Multiple Classifier Using Local Accuracy Estimates. IEEE Trans. on PAMI 19(4), 405–410 (1997)
18. Xu, L., et al.: Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition. IEEE Trans. on SMC 22(3), 418–435 (1992)

Processing of Missing Data in a Fuzzy System

Sylwia Pospiech-Kurkowska

Silesian University of Technology, Institute of Electronics, Akademicka 16, 44-101
Gliwice, Poland
sylwia.pospiech@polsl.pl

Summary. The more information is processed in a system the more likely is that some input values are missing. The paper describes (i) a method for managing the incomplete input data in a Mamdani fuzzy system and (ii) discusses the influence of inference interpretation on an efficiency of the fuzzy system operating on incomplete data. Two fuzzy models of missing information are discussed theoretically and then presented on an example of iris data set. Various interpretations of fuzzy rules and various types of membership function are examined in order to find a solution of fuzzy system that is more robust to missing data.

1 Introduction

The more we know about the patient the better is his/her diagnosis and treatment so usually in the medicine large amounts of information are collected before a decision is taken. However frequently not all the potentially significant information is available for a given patient. Reasons can be various: some examinations can be risky for the patient, too expensive to be performed routinely, or sometimes a patient deliberately withholds some information in questionnaire. Such incompleteness of input information is a problem a human being usually deals with very efficiently. On the contrary a fuzzy inference system where the knowledge is represented by rules (and the majority of the automatic reasoning methods), it is unable to produce a result, if the input data does not match a predefined set of attributes. How to deal with the problem? Elaborating a special set of rules for every possible case of missing data seems impractical, and even unachievable for exploding complexity in systems with many inputs. It would be better to make a fuzzy system with a given rule base able to process incomplete data sets. A solution proposed in this paper is to replace a missing datum by a fuzzy model of unknown value. Another question arises now: whether it is possible to minimize the deterioration of results for partially missing data by special choice of fuzzy operators interpreting the inference process.

The problem of dealing with missing information has already drawn the attention of researchers. Common probabilistic solution is to replace the missing value with its estimate obtained from the conditional probability distribution given the known features. In a fuzzy setting Berthold and Huber [1] describe a method for training a classifier with incomplete data. In normal operation of the classifier a missing value on input is assigned a degree of membership equal to

one. Gabrys [2] presents a generalized fuzzy min-max neural network that can deal with incomplete information both during the design and the normal operation. The missing values are represented as real valued intervals spanning the whole range of possible values. The result of the classification in this situation is rather a reduced number of alternative classifications than one class. In an Austrian fuzzy diagnosis support system for various fields of medicine [3] a user is allowed to specify some elements as 'unknown', but the internal representation is not explained. In the system for assessment of skeletal development of children [4, 5] from hand radiographs the ability to deal with incomplete data is achieved by two-step inference process and dividing the classifier into subclassifiers. Each region of interest in the hand has its classifier, that is activated only when parameters of the region are measured, independently from the others subclassifiers. A final classification is then obtained by a special 'averaging' aggregation of results from the previous step.

2 Processing of Incomplete Data in Theory

2.1 Fuzzy Inference System

Knowledge in a fuzzy inference system is represented by rules. Each rule in a Mamdani system with a multiple input and a single output has a form:

$$x_1 \text{ is } A_1 \wedge x_2 \text{ is } A_2 \wedge \dots \wedge x_k \text{ is } A_k \Rightarrow y \text{ is } B \quad (1)$$

Where

x_1, x_2, \dots, x_k - values of input variables

A_1, A_2, \dots, A_k - fuzzy sets for input variables

y - value of output variable

B - fuzzy sets of output variable

Fuzzy set A is described by its membership function μ_A with values in the range $[0,1]$. Statement $x_k \text{ is } A_k$ is interpreted as supremum of intersection of the fuzzy sets x_k and A_k . Connective \wedge (and) can be interpreted by minimum operator or by product operator. Value obtained from evaluation of the premise is called firing level of the rule. Implication \Rightarrow is interpreted by minimum operator or by product operator. If there are several rules, then the inference result is obtained by aggregation of outputs of all rules. Aggregation is usually performed by maximum operator or probabilistic or (probor). Whenever a crisp output is necessary, a defuzzification is applied.

2.2 Representation of Missing Data

Consider a fuzzy inference system with n input variables. Assume that a datum for k -th input is missing. How to execute the inference process in this case? Note that the only problem that needs a special solution is evaluation of premises where k -th variable is missing. Other effects for the inference process are caused by a way a modified firing level of the rule projects on the conclusion and the aggregation of rules.

‘Skip’ approach: if the k -th input is missing, then simply skip the statement $x_k \text{ is } A_k$ in the premise of the rule. Firing level of the rule in this case is calculated as:

$$\bigwedge_{\substack{1 \leq i \leq n \\ i \neq k}} \text{sup}(x_i \cap A_i) \tag{2}$$

and as from definition of the membership function:

$$0 \leq \text{sup}(x_i \cap A_i) \leq 1 \tag{3}$$

so from (2) and (3) it can be concluded:

$$\bigwedge_{\substack{1 \leq i \leq n \\ i \neq k}} \text{sup}(x_i \cap A_i) \geq \bigwedge_{1 \leq i \leq n} \text{sup}(x_i \cap A_i) \tag{4}$$

That means that the firing level of the rule with unknown k -th input is always greater or equal to the firing level of the rule with all known inputs. Statement (4) is true independent of the connective interpretation, however with product as \wedge , firing of the rule is more likely to be higher for missing data then for complete. Note that ‘skip’ representation is equivalent to assuming a degree of 1 for the statement $x_k \text{ is } A_k$, which is similar to approach proposed in [1].

‘Null’ approach: if the k -th input is missing, then substitute it by a ‘null’ fuzzy set representing unknown value. Since no information about x_k is available it seems reasonable to assume any value in the domain is equally possible. The ‘null’ fuzzy set representing unknown value for the k -th input variable has a membership function equal to constant $c = 0.5$ over the entire domain of the k -th variable (Fig. 1). The cutoff at $c = 0.5$ was chosen as representing the state between not being a member and being a member of a set. In this case the premise is evaluated:

$$\bigwedge_{1 \leq i \leq n} \text{sup}(x_i \cap A_i) = \bigwedge_{\substack{1 \leq i \leq n \\ i \neq k}} \text{sup}(x_i \cap A_i) + \text{sup}(x_k \cap A_k) \tag{5}$$

And as from definition of the membership function

$$0 \leq \text{sup}(x_i \cap A_i) \leq c \tag{6}$$

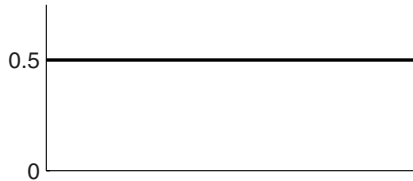


Fig. 1. In a ‘null’ approach the unknown input value is represented by a fuzzy set with a membership function equal to constant $c = 0.5$ over the entire domain of the input variable

If connective \wedge is interpreted by minimum, then we get:

$$\min(\min_{\substack{1 \leq i \leq n \\ i \neq k}}(\sup(x_i \cap A_i)), c) \leq c \tag{7}$$

That means the rule with missing data is not fired at higher level then for complete data and the firing level is cut off at $c = 0.5$. If connective \wedge is interpreted by product, then we get:

$$\text{prod}(\prod_{\substack{1 \leq i \leq n \\ i \neq k}}(\sup(x_i \cap A_i)), c) = c \cdot \prod_{\substack{1 \leq i \leq n \\ i \neq k}}(\sup(x_i \cap A_i)) \tag{8}$$

That means the firing level of the rule with missing data is reduced by a factor of $c = 0.5$ in comparison with the rule for complete data.

3 Dealing with Incomplete Data on Exaple of Iris Data

In this section we demonstrate how the proposed methods of managing missing data work and how the interpretation of inference influences the efficiency of the fuzzy classifier. Iris plant data are used for tests. This well known data set [6] published by R.A. Fisher in 1936 is often applied in tests of classifiers. The set counts 150 cases, for each case four features (sepal length, sepal width, petal length, petal width in cm) and a proper classification into one class (1: Iris versicolor, 2: iris setosa, 3: iris virginica).

3.1 Applied Fuzzy Inference Systems

Tests have been conducted on fuzzy systems with various types of membership functions and different interpretation of the inference process. However the basic structure of all systems is the same. Domain of each input variable is covered by three fuzzy sets representing values typical for every class. Three types of membership functions have been applied: gaussian (*gaus*), triangular (*tri1, tri2*) and trapezoidal (*trap*). The membership functions are generated automatically on the base of the data. Output space is divided into three fuzzy sets corresponding to classes. The same type of the membership function is used at output as for inputs. Parameters of the membership functions for each input are calculated in the similar way. The center of membership function is located at mean value of k -th feature for elements belonging to a given class. In trapezoid membership function the area at mean +/- standard deviation is equal 1. For Gaussian membership function the second parameter is the standard deviation of the feature in the class. The first and the last parameter of triangular and trapezoidal functions is placed where a membership function reaches zero. Their positions are chosen in order to obtain the membership 0.25 for the minimum and maximum in the class respectively (*tri2, trap*) or to obtain membership 0.5 at 25 and 75 percentile (*tri1*). Parameters of the membership functions can be found in the tab.3-6.

Table 1. The evaluation of the fuzzy system tr11 with the complete data ($x_1 = 6.20$, $x_2 = 2.90$, $x_3 = 4.30$, $x_4 = 1.30$, proper *class* = 2) and missing value for variable x_3 . Two interpretations of and are used minimum or product operator.

	Complete data				Missing x_3 - skip				Missing x_3 - null					
Rule	$x_i is A_{i,k}$				Firing		$x_i is A_{i,k}$		Firing		$x_i is A_{i,k}$		Firing	
$k \setminus i$	x_1	x_2	x_3	x_4	min	prod	x_3	min	prod	x_3	min	prod		
R_1	0	0.69	0	0	0	0	-	0	0	0.5	0	0		
R_2	0.84	0.86	0.97	0.95	0.84	0.66	-	0.84	0.68	0.5	0.50	0.34		
R_3	0.85	0.94	0.21	0.23	0.21	0.04	-	0.23	0.18	0.5	0.23	0.09		

A rule base is the same in every system and consists of three rules:

$$\begin{aligned}
 R1 &: x_1 is A_{1,1} \wedge x_2 is A_{2,1} \wedge x_3 is A_{3,1} \wedge x_4 is A_{4,1} \Rightarrow y is B_1 \\
 R2 &: x_1 is A_{1,2} \wedge x_2 is A_{2,2} \wedge x_3 is A_{3,2} \wedge x_4 is A_{4,2} \Rightarrow y is B_2 \\
 R3 &: x_1 is A_{1,3} \wedge x_2 is A_{2,3} \wedge x_3 is A_{3,3} \wedge x_4 is A_{4,4} \Rightarrow y is B_3
 \end{aligned}$$

The first index at A describes the input number, the second describes the class, and the index at B describes the class. Inference process is interpreted by minimum, maximum, product and probabilistic or operators as described in section 2.1. Two most popular methods for defuzzification are used: center of area (COA) and mean of maxima (MOM).

3.2 Example for a Single Case

This section demonstrates step by step how one missing input value influences the operation of the fuzzy system. A case where $x_1 = 6.2$, $x_2 = 2.9$, $x_3 = 4.3$, and $x_4 = 1.3$, proper *class* = 2 is considered. Tab.1 shows the evaluation of each statement $x_i is A_{i,k}$, and the firing for each rule for complete data and missing value for x_3 . Rule R_1 is not fired, rule R_2 is fired at lower level for null representation, at equal or slightly higher level for skip representation, rule R_3 is fired at higher level for both representations. The change in firing levels is more substantial for prod than for min. Fig.2 shows the output fuzzy sets for rules, the degree of membership for B_2 (proper class) is smaller and the degree of membership for B_3 higher for both representations of missing value, that means the fuzziness of the classification increases in comparison to the result for complete data, but the defuzzified values still point the proper class.

3.3 Results

Fuzzy systems based on product/probabilistic or operators showed to be more sensitive to missing data than systems based on minimum/maximum operators tab.2. However the deterioration of the classifier’s performance depends strongly on which variable is missing. Features x_3 and x_4 are much more crucial for

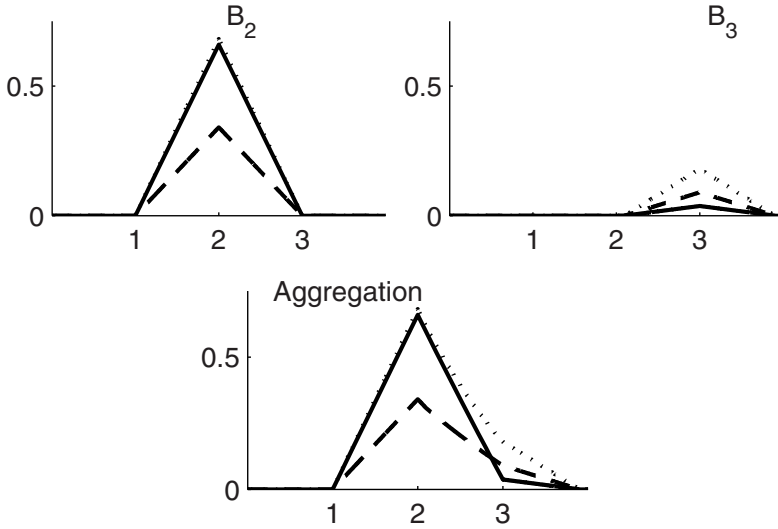


Fig. 2. Outputs of rules R_2 , R_3 and aggregation for complete data (rule R_1 is not fired, see tab.1) - solid line, for missing x_3 : skip representation - dotted line, null representation - dashed line (classifier tri1, implication prod). After MOM defuzzification 2 for all, COA: complete data 2.05, null 2.19, skip 2.18.

Table 2. The increase in number of wrong classifications respectively to the same fuzzy system with complete input data. A comparison of null and skip representations, min/max vs. prod/probor interpretation of inference.

	min/max				prod/probor											
	Skip		Null		Skip		Null									
var	x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4				
gaus	0	0	1	8	0	0	2	7	0	1	3	11	0	1	3	11
tri1	0	0	3	6	0	0	1	6	-1	-1	5	6	-1	-1	5	7
tri2	0	0	3	4	0	0	3	6	-3	-1	5	6	-2	-1	5	6
trap	0	0	1	6	-1	-1	1	6	-1	0	4	7	-1	0	4	7

classification then features x_1 and x_2 . Influence of membership function type is not considerable. The difference in efficiency of the classifiers with null and skip representation is not significant. This effect can be caused by the fact that in the considered rule base each input variable is present in premises of all rules that is a change in one rule conclusion is compensated by similar changes in the output of the other rules. Results for MOM defuzzification are similar as for COA.

Table 3. Gaussian membership function *gaus* - parameters

Variable	Class 1		Class 2		Class 3	
x_1	0.352	5.006	0.516	5.936	0.636	6.588
x_2	0.381	3.418	0.314	2.770	0.322	2.974
x_3	0.174	1.464	0.470	4.260	0.552	5.552
x_4	0.107	0.244	0.198	1.326	0.275	2.026
y	0.500	1.000	0.500	2.000	0.500	3.000

Table 4. Triangular membership function *tri1* - parameters

Variable	Class 1			Class 2			Class 3		
x_1	3.947	5.006	6.197	4.382	5.936	7.532	4.056	6.588	8.556
x_2	1.741	3.418	4.891	1.615	2.770	3.715	1.813	2.974	4.213
x_3	0.768	1.464	2.118	2.370	4.260	5.520	3.974	5.552	7.574
x_4	0.028	0.244	0.778	0.837	1.326	2.037	1.087	2.026	2.737
y	0.000	1.000	2.000	1.000	2.000	3.000	2.100	3.000	3.900

Table 5. Triangular membership function *tri2* - parameters

Variable	Class 1			Class 2			Class 3		
x_1	4.065	5.006	6.065	4.555	5.936	7.355	4.337	6.588	8.337
x_2	1.927	3.418	4.727	1.743	2.770	3.610	1.942	2.974	4.075
x_3	0.845	1.464	2.045	2.58	4.26	5.38	4.149	5.552	7.349
x_4	0.052	0.244	0.71866	0.891	1.326	1.958	1.191	2.026	2.658
y	0.000	1.000	2.000	1.000	2.000	3.000	2.100	3.000	3.900

Table 6. Trapezoidal membership function *trap* - parameters

Variable	Class 1				Class 2				Class 3			
x_1	4.133	4.800	5.200	6.000	4.667	5.600	6.300	7.233	4.467	6.200	6.900	8.233
x_2	2.033	3.100	3.700	4.633	1.833	2.500	3.000	3.533	2.000	2.800	3.200	4.000
x_3	0.867	1.400	1.600	2.000	2.667	4.000	4.600	5.267	4.300	5.100	5.900	7.233
x_4	0.067	0.200	0.300	0.700	0.933	1.200	1.500	1.900	1.267	1.800	2.300	2.567
y	0.000	1.000	1.000	2.000	1.000	2.000	2.000	3.000	2.200	3.000	3.000	3.800

4 Conclusions

A method for dealing with missing input values has been presented that enables a classical Mamdani fuzzy system to process incomplete data. No significant difference in overall classifier performance between the two proposed

representations of missing data was observed in tests for iris plant data. The inference interpreted by minimum/maximum operators has turned out to be more robust with respect to incomplete information than the inference based on product/probabilistic sum. Future investigations should include tests for more varied rule bases and non relational interpretations of implication.

References

1. Berthold, M.R., Huber, K.P.: Missing values and learning of fuzzy rules. *Int. Jour of Uncertainty, Fuzziness and Knowledge-based Systems* 6(2), 171–178 (1998)
2. Gabrys, B.: Neuro-fuzzy approach to processing inputs with missing values in pattern recognition problems. *Int. Journal of Appr. Reasoning* 30, 149–179 (2002)
3. Boegl, K., Adlassing, K.P., Hayashi, Y., Rothenfluh, T., Leitich, H.: Knowledge acquisition in the fuzzy knowledge representation framework of a medical consultation system. *Artificial Intelligence in Medicine* 30, 1–26 (2004)
4. Pospiech-Kurkowska, S., Gertych, A., Pietka, E.: Rozmyty system do szacowania rozwoju kostnego na podstawie rentgenogramow. In: *Proceedings of BIB Conf. Biocybernetyka i Inzynieria Biomedyczna*, pp. 1056–1061 (2005)
5. Pietka, E., Pospiech-Kurkowska, S., Gertych, A., Cao, F.: Integration of Computer assisted bone age assessment with clinical PACS. *Comp. Med. Imaging and Graphics* 12, 217–228 (2003)
6. Fisher, R.A.: Iris dataset: <ftp://www.ics.uci.edu/~mlearn/MLRepository.html>

Knowledge-Based Decision Hybrid System for the Doctor's Work Support

Zbigniew Buchalski

Institute of Computer Engineering, Control and Robotics, Wrocław University of Technology, Janiszewskiego str 11/17, 50-372 Wrocław, Poland
zbigniew.buchalski@pwr.wroc.pl

Summary. In the paper certain expert system concept for neediness of patients disease entity diagnoses was presented. Problem of knowledge acquisition to knowledge base and conclusion on that knowledge were realized. Testing results of HYBRIDEX expert system were presented as well as the effectiveness of that system usage as a doctor support in setting correct medical diagnose was estimated.

1 Introduction

Computer support of human activity takes place in various domain of science and technology [2, 3, 7]. One of domains, in which development was possible due to the usage of computers is artificial intelligence [4, 5]. Dynamic growth of science branches using artificial intelligence led to great interest of expert systems [2, 3, 4].

Deployment of expert systems in medicine can considerably increase effectiveness of doctor work. It's because of great knowledge, which general practitioner must know in order to make a correct diagnosis of chosen ailment case. The answer on question about usability and effectiveness of expert system depends on solution of knowledge acquisition into knowledge base problem and next on method of conclusion process based on stored knowledge.

The purpose of that paper is computer realization of knowledge acquisition into knowledge base and that knowledge elements conclusion process in certain expert system called HYBRIDEX. The basic task of the paper is estimation of effectiveness in expert system knowledge gathered in knowledge base management as a doctor support in correct medical diagnoses.

Knowledge base was built by expert in laryngology and contains rules and facts describing certain disease entities symptoms, that can be diagnose in hearing ailments (the scope of the expert system usage is hearing and balance diagnoses) [1]. The task of expert system is initialization of man-computer dialogue. The result of that dialogue is presentation of expert system advices and informations, which help in making correct diagnosis of certain patient.

2 Knowledge Representation in HYBRIDEX System

In the latest time new category of artificial intelligence tools appeared – hybrid systems. In general we can say, that they are connection of traditional expert systems, self-learning systems, neural networks and genetic algorithms.

Presented in the paper HYBRIDEX expert system is hybrid system, which allows alternative conclusion mechanism to rule based conclusion. It's made so, because usage of rules as an only tool can much diminish effectiveness of knowledge management.

The hybrid expert system model is presented in the following figure:

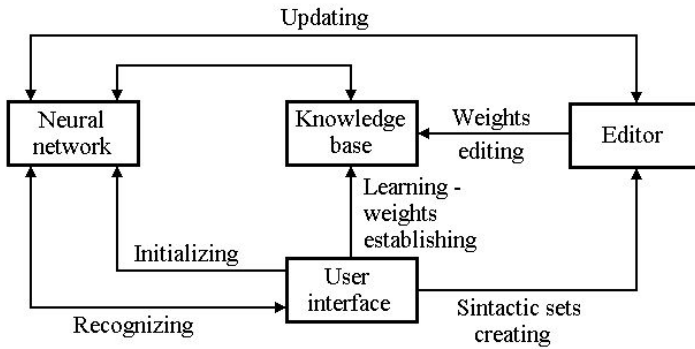


Fig. 1. Hybrid expert system model

Hybrid system application is especially useful in case, when knowledge necessary to conclusion process is difficult to formalization, uncertain or incomplete. In many disease entities providing of complete knowledge to system about disease causes, symptoms or ways of cure is in practice not possible.

Advantage of hybrid system construction is full cooperation of neural network and expert system, in which conclusion is based on rules and where problem can be broken down and solutions of subproblems thanks to full integration can be passed from one technique to another.

The easiest way of hybrid system creation is connection of traditional, based on rules conclusion methods with neural networks. On this ground symbolic transformation, which characterize traditional expert system is complementary to dispersed parallel transformation, specific for neural networks.

In the simplest case both types of transformation can be done independently in the same time. In such a solution one of environments is established as more important. This environment is responsible for tasks distribution into systems. The choice of system, which will solve task and assure the best solution depends on problem type.

HYBRIDEX system contains user interface, transforms data into proper coefficients, searches knowledge base. Next to that system activities, steering is passed to neural network in order to esteem, what are the most probable expert's report results.

It's well known, that in many knowledge domains creation of clear and explicit rules is not possible. On this ground neural network is suitable tool for that type fuzzy and inexplicit knowledge transformation. Expert system creates file, which contains coefficients as input data for neural network. Earlier neural network was learned with data provided by knowledge sources providing proved and sure knowledge. After neural knowledge calculations results from output are given back to expert system, which is responsible for theirs interpretation and displays final results.

The basic neural networks feature is their ability to based on previous illnesses diagnoses knowledge assimilation. It is not necessary to specify solutions of concrete problem, the only thing is to gather big and representative set of symptoms, which are sufficient to recognize certain disease entity. Neural networks can in natural way transform set of gathered by expert's knowledge symptoms into expected results, that means they can diagnose concrete disease entity.

Important neural network feature is results generalization ability – they can diagnose certain disease entity when not full ailment symptom list was provided by patient. Neural network is also able to use inconsistent or incomplete data, that means it can recognize correct disease entity even though, patient provided some symptoms specify for another ailment.

HYBRIDEX system was created with usage of PC-SHELL expert system framework, which allows construction of traditional expert system as well as hybrid system thanks to integration of neural network simulator with framework systems in knowledge representation and architecture level [6].

3 HYBRIDEX System Conclusion Mechanism

The conclusion process on HYBRIDEX system's knowledge base purpose is displaying to user all results, that fulfil provided at the beginning set of conditions in form of rules stored in knowledge base. Conclusions of described system activities are stored in computer memory. Untrue results are not stored and just ignored. Next conclusion process task is affirmation or deny hypothesis, which can be put due to beginning set of conditions.

In expert system presented in the paper backward conclusion, called also from hypothesis to conditions conclusion was applied. The conclusion process is presented in figure 2.

Backward conclusion starts with hypothesis and searches for arguments (proofs), which can prove or deny that hypothesis. Backward conclusion task in general is affirmation of main hypothesis on the ground of premises affirmation. If it's not sure that premise is true, it is treated as new hypothesis, which has to be proved. If as a result of that new rule, for which all premises are true is found, that rule's conclusion is also true. On the ground of that conclusion

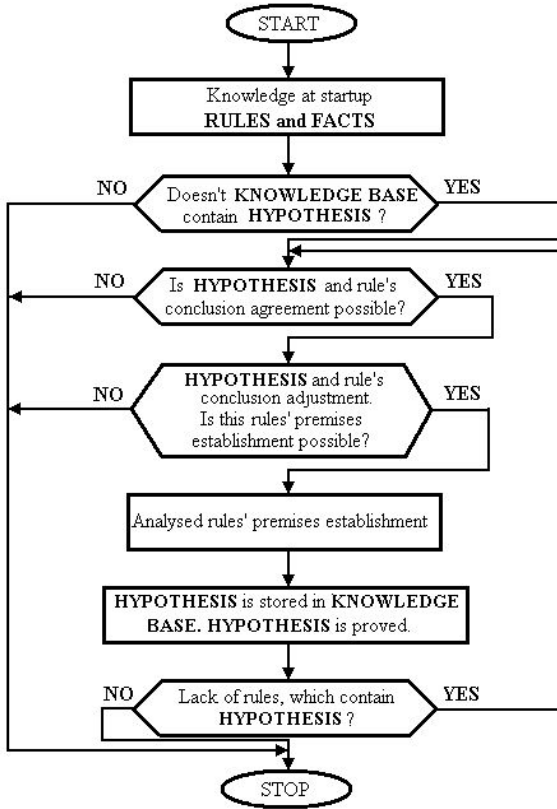


Fig. 2. Backward conclusion process model applied in HYBRIDEX system

the next rule, which premise wasn't previously known is being proved, a.s.o. Hypothesis, which was put is true only than, when all analysing premises are true.

Backward conclusion in HYBRIDEX expert system is in fact form of hypotheses verification. It comes from hypothesis (destination) through rules to facts. In practice it consist in hypothesis agreement with fact or rule. If the agreement is not possible, system is searching for next fact or rule and agreement operation is repeated. Agreement is comparison of matched hypothesis elements or rule's condition and fact or rule's conclusion. Agreement of objects and attributes takes place, when they're identical.

Important feature of HYBRIDEX expert system's algorithm is that in conclusion process chain of many rules invoking sequentially can be created.

Fundamental feature, which distinguishes backward conclusion and forward conclusion is smaller quantity of new facts generation and impossibility of many hypothesis affirmation. It is necessary to emphasize, that duration of backward conclusion is in many cases much shorter than in forward conclusion.

4 HYBRIDEX System Realization

HYBRIDEX expert system was constructed with SPHINX artificial intelligence package provided by AITECH company [6]. This package was chosen, because it allows programist easy interpretation and cooperation with neural network and provides easiness and simplicity of database creation, which is necessary in expert system using rule knowledge representation. Additional trump are defined user communication methods and ability to friendly and easy user-system communication interface.

Proposed in the paper HYBRIDEX hybrid expert system was created to support doctor work in disease entities recognition and diagnoses on the ground of two mechanisms:

- backward conclusion on the ground of rules and facts stored in knowledge base,
- solution specification on the ground of input data analysis by neural network.

User can choose conclusion mechanism on his own. In the disease entity recognition uncertainty second conclusion mechanism can be used to assure the most credible disease entity diagnose. After the most probable disease entity diagnose system advices to carry out proper medical examination to diminish likelihood of wrong diagnosis. In some cases system suggests hospital cure possibilities.

To make a diagnosis on the ground of input data (facts) multidimensional neural network created from connected perceptrons (neural cells) was used. It's normal procedure to organize neurons in layers. Input layer is specified with possible disease entities symptom quantity and output layer with possible to recognition disease quantity.

Basic neural networks feature is their ability to knowledge assimilation on the ground of provided examples. It's not necessary to specify certain problem solution, but only suitable big and representative set of samples. Neural networks can in natural way transform provided information and derive from it the most important elements.

Incontestable neural network advantage is their ability to incomplete and trouble data processing. In suitable example it can be shown, that on the ground of only few symptoms it is difficult to make not only correct diagnosis, but also any diagnosis. Conclusion on the ground of rules provides any results and in some cases even specialist can be surprised with diagnosis accuracy proposed by neural network. Moreover neural networks are esteemed for their quick work, which makes them valuable in real time systems.

Correct neuron calculation technique application often is much better than rule based conclusion methods. This difference can be well observed in work on great input and output parameters number.

The most important part of neural network implementation in describing project is its learning. This process has great impact on later system work correctness. Thanks to multidimensional network application learning process is controlled.

When network learning process is ended results usage is possible. In the proposed in the paper expert system neural network's work is absolutely hidden for user and depends on filling in form with symptoms, which is next sent to neural network for analysis and results receiving – diagnosis. After providing facts to input neural network undertakes analysis process and provides to output certain values, which after suitable interpretation are recognized as answer.

Diagnose based on first work mode, that means with rule knowledge representation subsystem application uses, what was meant backward conclusion. System user, who expects disease recognition at the beginning establishes first symptoms by writing them down in the table and next answers on question sequence provided by expert system. After such a dialogue with patient conclusion mechanism establishes additional facts.

Information gathered by system can be qualified as facts, when they are declared as disease entity symptoms in knowledge base. Conclusion mechanism searches for rules in which provided new fact appears. It doesn't ask twice the same questions to user when some facts were already established, but only analyse them and match them to rules stored in knowledge base. At the conclusion process beginning conclusion mechanism analyse large number of rules. When it has more facts these rules are eliminated, which don't contain selected facts or if specified fact excludes rule. During the whole conclusion process it tries to establish as much as possible facts to rule affirmation. To fulfil end of conclusion process condition all possible rules must be tested.

Conclusion system is supported by so called return mechanism, thanks to which finding of all accessible solutions is possible. That's why providing two diseases in the same time is nothing special. Some of diseases have similar symptoms in certain stage.

After facts are established expert system provides disease entity recognition. Establishing facts and hypothesis affirmation process investigation is also possible. Explanation about disease entity recognition can be provided after each diagnosis.

Conclusion based on HYBRIDEX expert system second mode work, that means with neural network application in recognition process, starts with providing disease symptoms to system. All provided by user symptoms are put into neural network input. On this ground it makes a diagnosis and proposes additional activities, which will prove certain disease entity recognition.

5 Conclusion

Presented in the paper HYBRIDEX expert system proves and shows abilities of artificial intelligence methods application in non computer science world. System purpose is doctor work support by providing certain possible solutions (theoretical disease entity diagnosis) and later activities for that solution affirmation. This system operation isn't ideal because of knowledge range necessary to gathered in order to complete knowledge base filling up.

HYBRIDEX expert system was designed as a system with two mechanisms, which recognise patient's disease entity. One mechanism doesn't provide proper diagnosis authenticity level. When two mechanisms are provided by system always alternative conclusion method can be used. When one mechanism doesn't suffice diagnosis credible enough it can be proved or denied by second one.

HYBRIDEX expert system test returned good results. System makes diagnosis correct not only in case, when rules are used, but also when neural network is applied. Conclusion on ground of rules didn't always return good diagnosis. It was caused by conclusion processing on incomplete data set. When only one rule's premise isn't proved, whole rule isn't proved and the possibility of lack of ailment recognition occurs. In the few cases recognition wasn't synonymous. Reason of that situation was either providing too many fact to knowledge base or providing by patient those symptoms, which are the same for related disease entities.

Neural network provided diagnosis in all cases. It should be emphasize, that neural network isn't main conclusion mechanism, but only additional tool in not credible results (one or more possible solutions) or when main conclusion mechanism working on rules doesn't find any solution. There was no such a situation, when neural network provided completely different solution than suggested by rules stored in knowledge base. In few cases, when conclusion mechanism working on rules didn't return any solution neural network provided it.

References

1. Becker, W., Naumann, H., Pfaltz, C.: Choroby uszu, nosa i gardla. Bel Corp, Warsaw (1999)
2. Buchalski, Z.: Knowledge management in the expert system for the decision-making process assist. In: Tadeusiewicz, R., et al. (eds.) Computer Methods and Systems. AGH University of Science and Technology, Cracow (2007)
3. Buchalski, Z.: Estimation of Expert System's Efficiency for the Computer Networks Design Assist. In: Grzech, A., et al. (eds.) Information Systems Architecture and Technology. Information Systems and Computer Communication Networks. Wrocław University of Technology, Wrocław (2007)
4. Buchalski, Z.: The role of Symbolic Representation of Natural Language Sentences in Knowledge Acquisition for Expert System. Polish Journal of Environmental Studies 16(4A) (2007)
5. Kazimierczak, J.: Computer Reasoning with Representation of Semantics of Natural Language Sentences. In: Proceedings of the 10th International Conference on System - Modelling - Control (2001)
6. Michalik, K.: Pakiet sztucznej inteligencji SPHINX. AITECH, Katowice (1998)
7. Zielinski, J.S. Inteligentne systemy w zarzadzaniu. Teoria i praktyka. PWN, Warsaw (2000)

Features for Text Comparison

Marek Krótkiewicz¹ and Krystian Wojtkiewicz²

¹ Institute of Mathematics and Informatics, University of Opole
mkrotki@math.uni.opole.pl

² Institute of Mathematics and Informatics, University of Opole
kwojt@math.uni.opole.pl

Summary. The main purpose of this paper is to deliver appropriate tool to find similarities between texts. The area of interest covers comparing large amount of different texts grouped in various areas of knowledge. Similarity is defined as distance between two texts and as this the measure may be calculated as the set of parameters based on features.

1 Introduction

This article focuses on the features useful for determining measure of similarity of texts. Comparison of texts in the meaning of determining the degree of similarity of texts is a complex task because of absence of unequivocal criteria of comparison. Texts may be compared from various points of view. The scrutiny of texts in terms of information conveyed in them is the most complicated undertaking. It requires both the syntactic analysis and the semantic one. Such approach requires, however, the use of the base of ontological knowledge by the system. Researches in this field have been conducted for a long time. Also the authors of this text have carried out the research topic connected with this problem. It does not change the fact that for a number of applications such a profound analysis is not necessary, in fact it is completely redundant. One should remember to choose an appropriate instrument for a given application.

The main area of studies described in this article is comparison of texts for determining similarity between them in morphological terms[1]. Here the basic morphological units are words, phrases and sentences. It should be emphasized, at the same time, that we do not take into account a grammar or semantics of utterances. Thanks to such an approach it is possible to achieve very efficient processing of texts, e.g. in order to determine whether a given text was written on the basis of another. Another application consists in determining an author of the text by means of recognition of objects with a learning set.

2 The General Description of the Problem

The aim of the authors' work was defining the features that can help to determine measures of similarity of texts. At the beginning we have to explain what the

notion of text means in this approach. In the most analytical sense, the text is a sequence of signs. The study of the text on a low level, however, does not allow to refer to its character, while the examination of its semantics is too complex for applications concerning defining measures of similarity of texts[2]. Therefore the authors decided to choose the middle approach, i.e. the treatment of the text as a sequence of words. The words has become the basic unit and the smallest element of the text.

Words as the basic units are defined as sequence of signs divided by empty signs. Phrases are sets of words occurring in uninterrupted flow. Sentences are sequences of signs ending with a sign of the end of a sentence. In consequence, the text consists of a sequence of sentences, sentences consist of phrases, phrases of words, and words are sequences of signs.

It is assumed that there are two texts. The aim is to define such features that would enable us to determine similarities of these texts by means of certain measures. The measures of similarity depend strictly on the aim of comparison and on the character of the texts in question, which will be discussed in the further part of this article.

The topic has been divided into three levels: words, phrases and sentences[3]. For each of these levels a group of features has been defined.

3 The Description of the Defined Features

A part of these features is dependent on parameters. So that it is possible to recognize texts and appoint the measures of their similarity for various applications without necessity of creating complicated, logical and computational, formulae. The parameters of the features:

LPS (Polish abbreviation) / **NRW** (English abbreviation) - minimal number of repeated words,

LSZ/NWS - the minimal number of words in sentences,

ULPS/PNRW - the minimal participation of the number of repeated words,

DF/LPh - the minimal length of a phrase,

MINDS/MINLS - the minimal length of an analyzed sentence.

The meaning of these parameters will be explained in a more precise manner while describing particular features.

Comparison on the level of words

Words constitute the lowest level. This set of features does not refer to environment, i.e. it is a comparison completely alienated from any context. It is coincidence of words in the texts examined that is analyzed here. Four features have been developed.

S_LPS/W_NRW - The number of repetitions of words in documents. The repetition of a word in a given sentence of the document is counted separately. If a given word of the source document is repeated several times in a given sentence or in various sentences, the repetitions in the document compared are counted

independently for each occurrence. For example in the texts given below: the parameter `S_LPS/W_NRW` will acquire value 9. Words *bbb* and *ccc* generate together three repetitions. The word *aaa* in the first sentence of the document

<p>T1.TXT: (<i>source document</i>) aaa bbb aaa aaa. ccc ddd eee.</p>	<p>T2.TXT: (<i>document compared</i>) aaa bbb ccc aaa aaa. bbb aaa aaa.</p>
---	---

T1.TXT appears three times and for each of these one repetition will be found in the first sentence and independently in the second one of the document T2.TXT, which together gives 6 repetitions. Thus the number of repetitions is 9. One has to notice that in the document T2.TXT in the first sentence the word *aaa* occurs three times and in the second one two times. However within one sentence of the document compared all cases of the word are counted as one occurrence.

S_ULPS/W_PNRW - The percentage participation of the number of repetitions of words in documents (`S_LPS/W_NRW`) in relation to the total number of words in the source document. In the example given above the source document has 7 words, so that this parameter will acquire the value of 128.6%.

S_LPULPS (ULPS)/W_NCPNRW (PNRW) - The number of comparisons of sentences, in which the percentage of repeated words for sentences compared equalled at least `ULPS/PNRW`. For the example given above the value of the parameter `ULPS/PNRW` was calculated as 35%. The total number of comparisons of sentences is 4, since each of the documents has two sentences. The first sentence of the source file (T1.TXT) as compared with the sentences of the file T2.TXT fulfills this condition twice, because in each case more than 35% of words is repeated in this sentence. In the case of the second sentence at best 1/3 of the words is repeated in T2.TXT. As a result the value of the feature `S_LPULPS/W_NCPNRW` is 2.

S_ULPULPS (ULPS)/W_PNCPNRW (PNRW) - The participation of the number of comparisons of sentences, where the percentage of repeated words for compared sentences was at least `ULPS/PNRW` in relation to the total number of comparisons. For the example in question the parameter `ULPS/PNRW` was 2, and the total number of comparisons was 4, so that the parameter `S_ULPULPS/W_PNCPNRW` was 50%.

Comparison on the level of phrases:

Phrase is a sequence of words, the length of which is not smaller than 2 words. Thanks to the parameter `LPh` for some features this value can be increased. There are six features prepared for phrases.

F_LPF/Ph_NRPh - The total number of repeated phrases. Algorithm of the search of repetitions was prepared in such a way as to avoid the repeated counting of the same phrases. As a result, the number calculated is modified by the number of repetitions shorter than the phrase of a given length. It means that, if there is a phrase consisting of 4 words, after the calculation has been completed, the total number of repeated phrases should be decreased by extracting 2. It is a consequence of the fact that a repeated phrase with minimal length of 4 words

means that there had to appear a phrase consisting of 3 and 2 words. Moreover, in order to avoid detection of nested repetitions of phrases, the next phrase in the sentence of the document compared is searched for beginning with the next word occurring after the last phrase found. The number of occurrences of phrases in the file compared, for the example given above, is 3.

F_WZLPF/Ph_ReSNRPh - The relative number of repeated phrases in relation to the number of comparison of sentences, i.e. the product of the number of sentences in both documents. Because of relatively large number of comparisons of sentences in relation to repetitions of phrases, this feature is multiplied by 1000. In the example analyzed above the value of this feature is $1000 \cdot 3/4 = 750$.

F_WSLPF/Ph_ReWNRPh - The relative number of repeated phrases in relation to the number of comparisons of words, i.e. the product of the number of words in both documents. Because of relatively large number of comparisons of words in relation to repetitions of phrases, this feature is multiplied by 1000000. In the example analyzed above the value of this feature is $1000000 \cdot 3/14 = 214285.7$.

The following three features are analogous to the previous ones. The only difference is that for calculations one takes phrases the length of which is equal at least to the parameter LPh.

F_LPFDf (DF)/Ph_NRPhLPh - The number of repetitions of phrases for assigned minimal length of phrase LPh. For the example analyzed above the value of this feature is 1.

F_WLPDFLZ (DF)/Ph_ReNRLPhNS (LPh) - The relative number of repetitions of phrases (multiplied by 1000), for assigned minimal length of phrase (LPh), in relation to the number of comparisons of sentences, i.e. the product of the number of sentences in both documents. For the example analyzed above the value of this feature is 250.

F_WLPDFLS (DF)/Ph_ReNRLPhNS (LPh) - The relative number of repetitions of phrases (multiplied by 1000000), for assigned minimal length of phrase (LPh), in relation to the number of comparisons of words, i.e. the product of the number of words in both documents. For the example analyzed above the value of this feature is 71428.6.

Comparison on the level of sentences

The following features concern the number of sentences repeated as a whole in both documents or when respective (parameter PSL/NRW) number of words is repeated in them. There are nine features prepared for sentences. The examples of texts to be analyzed have the following form: Parameters acquire the

Z1.TXT: (<i>source document</i>)	Z2.TXT: (<i>compared document</i>)
aaa bbb aaa aaa.	aaa bbb aaa aaa.
ccc ddd xxx eee yyy.	xxx yyy zzz.
fff ggg hhh iii.	jjj kkk.
jjj kkk.	

following value: $LSZ/NWS=3$, $LPS/NRW=2$. The group of six features given below concerns the situation when sentences are totally overlapping.

Z_LZ/S_NS - The total number of sentences repeated as a whole. For the example analyzed above the value of this feature is 2.

Z_ULZ/S_PNS - The percentage of the number of sentences in relation to the total number of sentences of the document. For the example analyzed above the value of this feature is 50%, since the number of repetitions of full sentences is 2, and the total number of sentences of the source document is 4.

Z_LS/S_NW - The sum of words repeated in sentences. For the example analyzed above the value of this feature is 6(4 words from the first sentence and 2 words from the fourth one).

Z_ULS/S_PNW - The percentage participation of the number of words in repeated sentences in relation to the total number of the of words in the document. For the example given above the value of this feature is 40%.

Z_LZLS (LSZ)/S_NSNW (NWS) - The number of repeated sentences longer than or equal to LSZ. For the example analyzed above only one sentence (the first one) is repeated and at the same time it consists of at least 3 words.

Z_ULZLS (LSZ)/S_PNSNW (NWS) - The percentage participation of the number of repeated sentences longer than or equal to LSZ in relation to the total number of sentences in the document. For the example given above the value of this feature is 25%.

Z_LPS (LPS)/S_NRW (NRW) - The number of sentences with repeated words, for the minimal number of repeated words LPS/NRW within one sentence. For the example given above the value of this feature is 3, since sentences no. 1, 2 and 4 fulfil this criterion.

Z_WZLPS (LPS)/S_ReSNRW (NRS) - The relative number of sentences with repeated words, for the minimal number of repeated words LPS/NRW within one sentence. For the example given above the value of this feature is 75%.

Z_WPZLPS (LPS)/S_ReRSNRW (NRW) - The relative number of sentences with repeated words, for the minimal number of repeated words LPS/NRW within one sentence, in relation to the total number of comparisons of sentences multiplied by 1000000 for legibility of this feature. For the example analyzed above the value of this feature is 250000.

4 Application

Due to the fact that there 19 features accessible, it is possible to create measures of similarity of very specific properties in terms of detection of the overlapping of texts. It is known that authors of texts possess specific inclinations for constructing their utterances in a certain way. One may define it, in a certain sense, as the style of writing.

The features concerning words and phrases may turn out to be most effective in this regard. Frequent repetition of words in two documents may point to one author. Obviously the focus on a single feature brings little effectiveness. When

similarity occurs simultaneously on the level of both words and entire phrases, one may suspect that it does not have to be an accidental coincidence. The features based on the relative values are essential here. They allow to refer particular values e.g. to the number of sentences in the text, comparisons of words or sentences. If comparison of two texts will show that the feature $Z_ULZ/S_PNS = 10\%$ and $Z_WZLPS(4)/S_ReSNRW(4)=30\%$, which means that 10% of sentences was precisely rewritten and 30% of sentences have at least 4 repeated words, one may suppose that the documents examined have common roots.

Correct description of parameters is essential for programs for defining the degree of textual similarity in terms of plagiarism. A special importance possesses the parameter $MINDS/MINLS$, which determines the minimal number of words in the sentence to make this sentence an object of analysis. It allows to filter out short sentences which obscure the image because of their insignificance from the point of view of communicated information. Moreover they may be repeated in texts as typical formulas which have nothing to do with plagiarism. The study of various texts show that this parameter should not be smaller than 3. It influences also the time of processing and can shorten it by several percent. The parameter DF/LPH defining the minimal length of phrase also contributes to filtering phrases that are too short to be considered as really borrowed ones.

Analyzing features in terms of proposed application we have to discuss the interpretation of some of them. The feature Z_LZ/S_NW describes the total number of overlapping sentences. It is a very strong feature. In texts possessing over 500 sentences and the parameter $MINDS/MINLS=3$, this feature equals zero if the texts are really written independently. Obviously, it is not a rule. However, the repetition of full sentences in two texts is a very serious signal. In practice, rarely does plagiarism consist in copying full sentences, therefore various, more precise features have been prepared. Derivative of Z_LZ/S_NW ? is the feature Z_ULZ/S_PNS which refers the number of full repeated sentences to the number of sentences of the analyzed text. It is obvious that there is a significant difference between repetition of 10 sentences in a text consisting of 200 sentences and the same repetition in a text of 2000 sentences, where the probability of an accidental repetition is much bigger. This feature can be applied as the measure of the level of copying one text to another, in the case of copying full sentences, i.e. an activity widely known as "copy-paste". Information complementary to the features Z_LZ/S_NS and Z_ULZ/S_PNS are values of the Z_LS/S_NW and Z_ULS/S_PNW . The Feature Z_LS/S_NW defines summary number of words found in the sentences classified during calculation of the feature Z_LZ/S_NS . In other words this feature is connected with the same "copy-paste" symptom, however in this case it is expressed in the form of the number of words and not sentences. Thanks to it is more precise as it is more sensitive to the length of the sentences copied. The feature Z_ULS/S_PNS has analogous interpretation, however it is expressed in a relative way related to the total number of words in the source document. The features Z_LZLS/S_NSNW and Z_ULZLS/S_PNSNW have analogous interpretation to Z_LZ/S_NS and Z_ULZ/S_PNS , at the same time they are capable of filtering out sentences which are too short.

The features S_LPS/W_NRW and S_ULPS/W_PNRW usually have high values in the case when documents compared concern the same thematic areas. As a result they should be used with caution and never in isolation.

A much more detailed analysis is provided the features connected with phrases. Our research has shown that the feature F_LPFDF/Ph_NRPhLPh, which determines the number of repeated phrases with an additional condition that the phrase cannot be shorter than the value of the parameter DF/LPh. The results show that the parameter DF/LPh should remain within the limit of 3-5 depending on the kind of the text. The feature F_WLPDFLZ/Ph_ReNRLPhNS constitutes the reference of the previous feature to the product of the number of sentences in both texts. Because of this, similarly to previous cases, it is possible to compare the values of the features for the texts of various lengths.

Requirements concerning algorithms

Because of the range of applications, directed at the screening examinations researches, the features described above and algorithm of their calculation were optimized for the purpose of efficiency. The computational complexity of calculating the features is great and non-linear. Each sentence of one document must be compared with each sentence of the other. All the features are calculated during a single course of all the sentences, which has a critical significance for efficiency. For each pair of sentences all the words of the first sentence are to be compared with the words of the second one. As a result, the essential indicators are:

- the time of comparison of compared documents in relation to the number of comparisons of sentences and
- the time of comparison of documents in relation to the number of comparisons of words.

It should be emphasized that the deepest resident fragment of the algorithm of the analysis of texts is the comparison of phrases. It is a consequence of the fact that for each coincidence of two words in sentences it should be determined if a given coincidence is or is not the beginning of the coincidence of the sequence of words, i.e. phrases. Efficiency research has been conducted. The length of the source text was 9543 words (772 sentences). The average speed of the processing of the text is 848 thousand comparisons of sentences per second (Table 1). The research was conducted with the use of a computer equipped with the processor Intel Pentium®D 3.0 GHz.

Table 1. Benchmark for sentence comparison

Comparison parameters	File no. 1	File no. 2	File no. 3	File no. 4
Sentences	377	1173	2016	2874
Sentences comparisons	272 194	846 906	1 455 552	2 075 028
Speed of sentences comparisons (thousands/s)	916.5	808.9	817.3	851.1

5 Conclusion

The method of quick comparison of texts presented here was optimized in terms of its employment in the wide-range screen examinations. Similarity is defined as distance between two texts and in this case the measure may be calculated as the set of parameters based on features. The further tasks will include implementation of algorithms based on semantic analysis in connection with the system of artificial intelligence in progress [4].

References

1. Salton, G.: Automatic text processing: the transformation, analysis and retrieval of information by computer. Addison-Wesley, Reading, Massachusetts (1988)
2. Tam, G.K.T.: Formal Concept Analysis and Text Similarity, Computer Science and Software Engineering, Monash University (January 2004)
3. Metzler, D., Dumais, S., Meek, C.: Similarity Measures for Short Segments of Text. In: Avances in Information Retrival. Lectures Notes in Computer Science, pp. 16–27. Springer, Heidelberg (2007)
4. Krótkiewicz, M., Wojtkiewicz, K.: Conceptual Ontological Object Knowledge Base and Language. In: 4th International Conference on Computer Recognition Systems - Advances in Soft Computing, pp. 227–234. Springer, Heidelberg (2005)

Possibility of Use a Fuzzy Loss Function in Medical Diagnostics

Robert Burduk

Chair of Systems and Computer Networks, Wrocław University of Technology,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
robert.burduk@pwr.wroc.pl

Summary. An application of a two-stage classifier to the prognosis of sacroileitis is presented in the paper. The method of classification is based on a decision tree scheme. A k -nearest neighbors is applied in pattern recognition task. In this model of classification a fuzzy loss function is used. The efficiency of this algorithm is compared with the algorithm based on zero-one loss function. In this paper also influence of choice of parameter λ in selected comparison fuzzy number method on classification results are presented.

1 Introduction

In many pattern recognition problems the loss (utility) evaluation can be quite imprecise [1]. For this reason, several studies have previously described decision problems in which values assessed to the consequences of decision are assumed to be fuzzy [2, 6, 7]. The class of the fuzzy-valued loss functions is definitely much wider than the class of the real-valued ones. This fact reflects the richness of the fuzzy expected loss approach to describe the consequences of incorrect classification as opposed to the real-valued approach. This paper deals with a recognition problem, where - assuming a probabilistic model - values of a loss function are assumed to be fuzzy numbers. We will also consider the so-called Bayesian hierarchical classifier [8]. In this recognition problem the decision as to the membership of an object in a given class is not a single activity, but is the result of a more or less complex decision process. This model has been formulated so that, on the one hand, the existence of fuzzy loss function representing the preference pattern of the decision maker can be established; while, on the other hand, a priori probabilities of classes and class-conditional probability density functions can be given in learning set.

In the paper [9] the Bayesian hierarchical classifier is presented. The synthesis of multistage classifier is a complex problem. It involves specifying the following components:

- the decision logic, i.e. hierarchical ordering of classes,
- feature used at each stage of decision,
- the decision rules (strategy) for performing the classification.

This paper is devoted only to the last problem. This means that we will deal only with the presentation of decision algorithms, assuming that both the tree skeleton and the feature used at each non-terminal node have been specified.

2 Hierarchical Classifier

In the paper [9] the Bayesian hierarchical classifier is presented. The synthesis of multistage classifier is a complex problem. It involves specification of the following components:

- the decision logic, i.e. hierarchical ordering of classes,
- feature used at each stage of decision,
- the decision rules (strategy) for performing the classification.

First two problems for prognosis of sacroileitis was presented in [5]. This paper describe only the last problem.

The procedure in the Bayesian hierarchical classifier consist of the following sequences of activities. At the first stage, there are measured some specific features x_0 . They are chosen from among all accessible features x , which describe the pattern that will be classified. These data constitute a basis for making a decision i_1 . This decision, being the result of recognition at the first stage, defines a certain subset in the set of all classes and simultaneously indicates features x_{i_1} (from among x) which should be measured in order to make a decision at the next stage. Now at the second stage, features x_{i_1} are measured, which together with i_1 are a basis for making the next decision x_2 . This decision – like i_1 – indicates features x_{i_2} necessary to make the next decision (at the third stage) and – again as at the previous stage – defines a certain subset of classes, not in the set of all classes, however, but in the subset indicated by the decision i_1 , and so one. The whole procedure ends at the last N -th stage, where the decision made i_N indicates a single class, which is the final result of multistage recognition. Thus multistage recognition means in medical diagnosis a successive narrowing of the set of potential diseases from stage to stage, down to disease unit, with symptoms which should be measured to make the next diagnosis more precise being simultaneously indicated at every stage.

3 Medical Description of the Problem

Sacroileitis (SI-itis) belongs to the group of rheumatoid arthritis [5]. SI-itis is chronic, mostly progressive inflammatory process, embracing lower back-hip joints, small spine joints, fibrous rings and lower back ligaments, leading to their gradual stiffeners. The reason of disease is not well-known, one can take into account the possibility of participation of infectious and hereditary factor. For participation of microbes speaks considerable frequency of contagions, embracing especially ureter in the period previous to symptoms of the disease. Genetic predisposition to falling ill with SI-itis is confirmed by its occurrence in ill families at about 20 % men and 8 % women of relatives of the first degree. It has been

showed, that antigen HLA B27 is present at 90-96 % of the ill in comparison to 5-14 % in the population of healthy testing persons. From among persons with positive antigen HLA B27 morbidity on sacroileitis comes up to only 1 %. It is considered, that the presence of the antigen testifies the genetic predisposition to falling ill under influence of environmental factor, probably infectious.

Correct diagnosis makes long time of development of disease, more difficult and large probability of first symptoms for different diseases. The period after which we have final disease results reaches for 5 to 10 years. Properly quick diagnosis of disease is important, so that correct treatments already in initial phase of occurrence of SI-it is undertaken. In such situation the attempt the computer-aided prognosis of the development of sacroileitis is well founded.

4 Data Description

Clinical material derive from the Research Institute of Rheumatic Diseases in Piestany. 84 patients with SI-it is were examined. Each case record was provided with final diagnosis made after 5 years.

In problem of multistage classifier following diseases and groups of diseases of rheumatoid type has been defined as classes:

- ankylosing spondylitis (definitive) (1),
- ankylosing spondylitis (probable) (2),
- sacroileitis persistens (3),
- others (4).

In group of others are following diseases: arthritis periferal, sacroileitis present, arthritis periferal, sacroileitis absent, Reiters syndrome, coxopathia only.

The multistage classifier for the prognosis of SI-it is development has been proposed in [10]. In this paper the hierarchial of disease units (the decision logic) and symptoms (features) used at each stage of diagnosis are presented. A two-level decision logic is presented in Fig. 1, where following natural number are described above. The interior nodes (0), (5) and (6) are described in turn sacroileitis, ankylosing spondylitis (Bechtterew disease) and other rheumatoid diseases. Table 1 list the features selected at each interior node of the two-stage classifier.

Table 1. List of the used syndromes

Node		
0	5	6
Stiffness of back	Stiffness of back	Wheel pain
Pain of S-I joints	Reiter's syn. in history	Ischalgia irradiating to knee
Wheel pain	Wheel pain	ASO
Tenderness of wheels	Lumbar pain	Tenderness of wheels
Reiter's syndrome in history	AS in family history	AS in family history

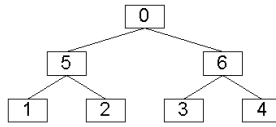


Fig. 1. Decision tree for the prognosis of SI-itis

5 Description of the Method

For the prognosis of SI-itis development, a dependent on the node of the decision tree fuzzy loss function was used. This loss function means that the loss depends on the node of the decision tree at which misclassification is made and allows imprecise, linguistic description of loss. We assume following linguistic description *small, medium, big* for loss function connected with suitable internal nodes (5), (6) and (0). In Fig. 2 are presented two cases of fuzzy loss functions used in the next section, which are represented by triangular fuzzy numbers.

Decision rule at each interior node are based on *k*-nearest neighbors (*k*-NN) rule for fuzzy loss function [3]:

$$\arg \max_{k \in M^{i_n}} \left\{ (\tilde{L}_{i_n} - \tilde{L}_k) k_l^{i_n} + \sum_{j_N \in M^{j_N-1}} \tilde{L}_{j_N-1} q^*(j_N/k, j_N) p(j_N) k_{j_N}^{i_n} \right\}. \quad (1)$$

For ranking fuzzy numbers we have selected the subjective method stated by Campos and González [4]. This method is based on the λ -average valued of a fuzzy number, which is defined for $\tilde{A} \in \mathcal{F}_c(\mathbb{R})$ as the real number given by

$$V_S^\lambda(\tilde{A}) = \int_0^1 [\lambda a_{\alpha 2} + (1 - \lambda) a_{\alpha 1}] dS(\alpha) \quad (2)$$

where $\tilde{A}_\alpha = [a_{\alpha 1}, a_{\alpha 2}]$, $\lambda \in [0, 1]$ and S being an additive measure on $Y \subset [0, 1]$.

The parameter λ is a subjective degree of optimism-pessimism. In a loss context, $\lambda = 0$ reflect the highest optimism and $\lambda = 1$ reflect the highest pessimism. Then, the λ -ranking method to compare fuzzy numbers in $\mathcal{F}_c(\mathbb{R})$ is given by

$$\tilde{A} \succeq \tilde{B} \Leftrightarrow V_S^\lambda(\tilde{A}) \geq V_S^\lambda(\tilde{B}). \quad (3)$$

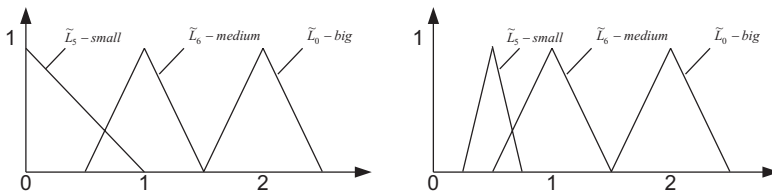


Fig. 2. Two cases of fuzzy loss function for the prognosis of SI-itis

6 Results of Recognition Algorithm

For the two-stage prognosis of the SI-itis development algorithm (1) was used for $k = 3$. The probability of correct classification $\hat{q}(i_n/i_k, i_n)$ was estimated by *leave-one-out* method. Results of frequency of correct classification are presented in Fig. 3.

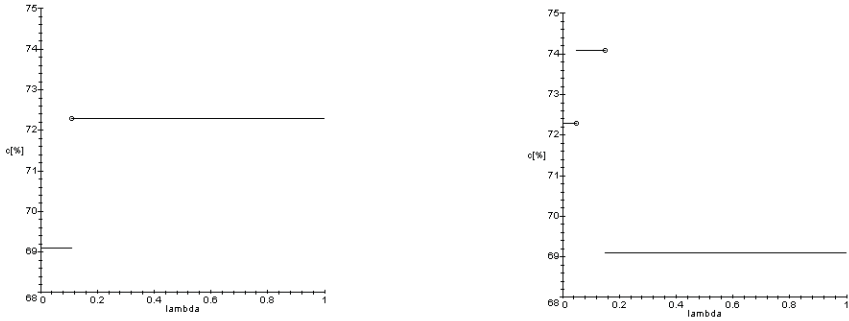


Fig. 3. Frequency [%] of correct classification

For zero-one loss function the correct classification is equal 69.1%. For selected cases of fuzzy loss function and for one here or there values of parameter λ we obtain better results of recognizing than for zero-one loss function.

7 Conclusion

This paper describe the multistage recognition task to the prognosis of the SI-itis development. Received results seem to confirm usefulness of multistage recognition algorithms with fuzzy loss function in the computer-aided medical diagnosis. Probably for the longer learning set the results of classification will be better. It is necessary to make an investigation for the other (parametric or nonparametric) estimators of conditional probability density functions.

References

1. Berger, J.: Statistical Decision Theory and Bayesian Analysis. Springer, New York (1993)
2. Baas, S., Kwakernaak, H.: Rating and Ranking of Multi-Aspect Alternatives Using Fuzzy Sets. *Automatica* 13, 47–58 (1997)
3. Burduk, R.: Computer Algorithms of Multistage Classifier with the Probabilistic-Fuzzy Model and Their Application in Medicine. PhD Thesis, Wrocław TU (2003)
4. Campos, L.M., González, A.: A Subjective Approach for Ranking Fuzzy Numbers. *Fuzzy Sets and Systems* 29, 145–153 (1989)
5. Meske, S., Lamparter-Lang, R.: Rheumatoid Arthritis, PZWL, Warsaw (in Polish) (1997)

6. Jain, R.: Decision-Making in the Presence of Fuzzy Variables. *IEEE Trans. Sys. Man and Cyber.* 6, 698–703 (1976)
7. Viertl, R.: *Statistical Methods for Non-Precise Data*. CRC Press, Boca Raton (1996)
8. Kurzyński, M.: Decision Rules for a Hierarchical Classifier. *Pat. Rec. Let.* 1, 305–310 (1983)
9. Kurzyński, M.: On the Multistage Bayes Classifier. *Pattern Recognition* 21, 355–365 (1988)
10. Kurzyński, M., Masaryk, P., Svec, V.: Multistage Recognition System Applied to the Computer-Aided Diagnosis of Some Rheumatics Diseases, *Systems Science, Wrocław* 15, 59–66 (1989)

An Application of a Generalized Additive Model for an Identification of a Nonlinear Relation between a Course of Menstrual Cycles and a Risk of Endometrioid Cysts

Dariusz Radomski¹, Zbigniew Lewandowski², and Piotr I. Roszkowski³

¹ Division of Nuclear and Medical Electronics, Institute of Radioelectronics, Nowowiejska 15/19. 00-665 Warsaw, Poland

D.Radomski@ire.pw.edu.pl

² Department of Epidemiology Medical, University of Warsaw, Oczki 3, 02-007 Warsaw, Poland

³ Department of Obstetrics and Gynecology, Medical University of Warsaw. Karowa 2, 00-315 Warsaw, Poland

Summary. Standard methods used for an identification of risk factors are based on logistic regression models. These models disabled to assessment a nonlinearity between a study factors and a disease occurrence. This paper presents an application of generalized additive models for modeling of reproductive risk factors associated with endometrioid cysts. Moreover theoretical similarity and differences between generalized additive models and neural networks was discussed. The obtained results enabled to propose a new etiological aspect for endometrioid cysts.

1 Introduction

An identification of risk factors associated with a disease is one of most frequent problems in epidemiology. Classically, such identification is performed on the base of a statistical inference about an odds ratio which is estimated by standard logistic regression models. Usually for this purpose continuous variables are arbitrarily discretizing and incorporated into model as dummy variables. This method allows for testing a linear trend between a discretized factor and a risk of a disease occurrence applying for example a Chi-square test for a linear trend [1]. However, this manner leads to significant loss of information accumulated in medical observations as well as increases measuring errors [2].

From etiological and pathophysiological points of view there is necessity to know a possible nonlinear dose-effect relationship between a tested risk factor and a probability of a disease occurrence. A large spatial-temporal averaging of analyzed processes based on epidemiological data causes that a possible form of nonlinearity is usually unknown explicite. Therefore, a nonparametric regression model for binary outcomes should be used.

In literature there are two main class of regression models devoted to this purpose. One of them are neural networks which according to Kolmogorov theorem and Hornik *et al.* theorem are universal approximators of any continuous

functions defined on a close set [3]. However, there are some problems associated with an application of neural networks for a statistical modeling of biomedical data. Firstly, there is a widely accepted procedure for selection of a neural network structure. Secondly, there is no an universal statistical test which enable to inference about statistical significant of an identified model. Moreover, relative complicated knowledge about neural network models limits their application in epidemiological analysis.

Alternative methods used for identification possible nonlinear relationship between risk factors and a disease occurrence are generalized additive models (GAMs). These models allowed use a well known likelihood ratio test for a non-linearity assessment.

The aim of the performed analysis was studying on a relationship between menstrual cycle characteristics and an occurrence of endometrioid cysts in women being in reproductive age. On the base of etiopathological data there are speculations that a length of a menstrual cycle as well as a length of a menstrual bleeding may play an important role in development of endometrioid cysts.

2 Modeling of Clinical Data

2.1 Clinical Data

Clinical data used for a nonlinearity assessment were obtained from the performed case-control study. The case group included 100 Polish patients (age range 19–42 years, mean 32 ± 8.72) diagnosed at the 2nd Department of Gynaecology and Obstetrics, Medical University of Warsaw for pelvic pains, dysmenorrhoea and/or infertility. Ovarian endometriosis (ovarian endometrial cysts) has been confirmed in these patients, both by laparoscopic and histopathological examination. The size of these cysts ranged from 2 to 8 cm and they were classified as stage III or IV of endometriosis according to the American Fertility Society (1985).

The control group was consisted of 200 age matched women drawn from the same population as the case group. Each woman of the control group was examined by gynecologists at most 6 months before the study. Potential endometriosis and malignant tumors of reproductive organs were excluded based on histories and clinical examinations. The age range of controls was 20–40 (mean 29.5 ± 9.76) years.

Data about courses of menstrual cycles were collected on the base of the specialized preparing questionnaire which was inquired by a gynecologist or a midwife.

2.2 Identification of the Relationship between a Menstrual Cycles Characteristics and a Risk of Endometrioid Cysts

A menstrual cycle course was characterized by the following variables: an age of menarche, an average length of a menstrual cycle, an average length of a

menstrual bleeding and the ratio of an average length of a menstrual bleeding to an average length of a menstrual cycle (denoting by τ).

Let r denote a continuous variable characterizing a menstrual cycle and D denotes a random binary variable representing an occurrence of endometrioid cysts. Then, a generalized additive model describing an investigated relationship has the following form:

$$D = \sum_{i=0}^I s_i(x_i) \tag{1}$$

where x_i is an element of a vector $\mathbf{X} = [r \ c_1 \ \dots \ c_I]^T$ which contains the analyzed variable and I confounders. s_i is a "smooth" function. The estimators $\hat{s}_i(x_i)$ are computed by maximization of the following likelihood function:

$$\ell(s_0, \dots, s_I) = l(z) - \frac{1}{2}Q(s_0, \dots, s_I) \tag{2}$$

where $z_k = \sum_{i=0}^I s_i(x_{ik})$ and $l(z)$ is a log-likelihood defined for exponential family of a probability density. The penalty function representing by $Q(s_0, \dots, s_I)$ has the following form:

$$Q(s_0, \dots, s_I) = \sum_{i=0}^I \lambda_i \int [s_i''(x)]^2 dx \tag{3}$$

Maximization of the formula (2) can be used by Newton–Raphston methods. A numerical version of this method leads to *local scoring algorithm* described by [4]. This algorithm was used to the identification of the studied models.

To assess a nonlinear versus a linear relationship between a study variable and a risk of endometrioid cyst two models were performed for each study variable. The first model had the form:

$$\mathcal{M}_1 : D = x_j + \sum_{\substack{i=0 \\ i \neq j}}^I s_i(x_i) \tag{4}$$

This model was linear towards to the study variable x_j . The second model was such as:

$$\mathcal{M}_2 : D = x_j + \sum_{i=0}^I s_i(x_i), \tag{5}$$

so it was smoothing nonlinear in relation to x_j . Let $\ell_1 \ell_2$ denote values of log-likelihood function for the first and second model, respectively. Then, the statistic $D = \ell_1 - \ell_2$ has Chi-square distribution with d degree of freedom under zero hypothesis which states that a study relationship is linear versus nonlinear. The degree of freedom is $d = n - 1 - tr(\hat{\mathbf{S}}_j^{-1})$, where n is a sample size and $tr(\hat{\mathbf{S}}_j^{-1})$ is a trace of the matrix which is inverse to a "smoothing" matrix $\hat{\mathbf{S}}_j = [\hat{s}_{k,l}]_{n \times n}$. Their elements are coefficients of the smoothing operator defined in [4]. The statistical inference was performed assuming $\alpha = 0.05$. Because this

model estimation is computation consuming covariates were omitted in the applied model. Thus fitted model selection was limited to finding the proper value of the smoothing parameter . This value was adjusted minimizing a mean square error between the model predicted values and the observed values. S-Plus 6.1 was used to perform statistical analysis.

3 Results

The application of the described method to the variables characterizing menstrual cycles enables to identify the nonlinear relationship between an average length of a menstrual cycle and a risk of endometrioid cysts. The relation between the analyzed risk and a length of a menstrual bleeding as well as τ parameter were linear. These relations were statistical significant. On the other hand, there was no statistical evidence for an association of the study risk with menarche age. The obtained results are presented in Fig. 1.

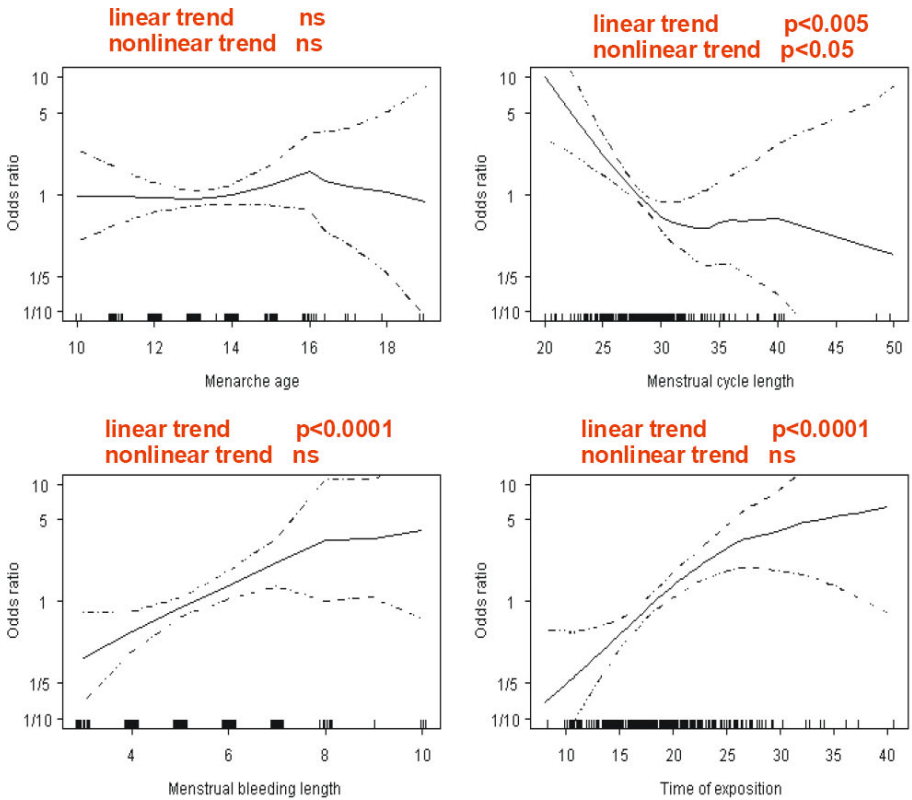


Fig. 1. The relationships between the tested variables characterized a menstrual cycle and a risk of endometrioid cyst occurrence

4 Discussion

Modeling of a relationship between a study variable and a risk of a disease is usually limited to an estimation of odds ratios. However, knowledge about risk factors associated with a given disease can be insufficient for understanding a disease etiology. Numerous previous studies showed that short menstrual cycles (length < 27 days) increased a risk of endometriosis [5]. These results were interpreted as a biological circumstance supported Sampson's hypothesis stating that a retrograde menstrual flow is an important etiological factor leading to dissemination of an ectopic endometrial tissue [6].

However, the application of the nonparametric nonlinear model showed that prolongation of menstrual cycles does not lead to decreasing of an endometrioid cyst occurrence. Moreover, women with oligomenorrhoea are characterized by slightly increasing of an endometrioid cyst risk. The obtained result suggests that there is another possible mechanism leading to development of an ovarian form (or each form) of endometriosis.

In our opinion, using of GAM models to a nonlinearity assessment has significant superiority in comparing to well known neural networks. Indeed, Anders *et al.* showed some applications of statistical tests to statistical inferences for neural networks [7]. Even, there are suggestions that well known Wald test could be used for this purpose. However, the Wald statistic contains a covariance matrix between model parameters (i.e. neural weights). This covariance matrix can be well estimated assuming that computed weights are global optimal. But it is well known that popular methods of neural networks learning do not lead to global identified models. It can be a serious limit for statistical inferences of neural networks. In opposite to them, GAMs enable to use well proven experimentally and robust statistical tests for nonlinearity assessment.

One of drawback of GAMs is a high dimension of an observation vector leads to hard computational problems which limit number of confounders used in the GAM. However, this problem can be solved in two ways. Hastie and Tibshirani showed a method of GAM identification assuming dependent observations [4]. It enables to use of propensity score matching for confounder controls [8]. Moreover, there is a parallel version of local scoring algorithm [9].

In conclusion, we state that an application of generalized additive models to epidemiological problems allows for exploring a new possible etiological mechanism. Nonlinearities identified by GAMs can be further modeled by parametric regression models to precisely investigate differences between stages or forms of a disease.

References

1. Agresti, A.: Categorical Data Analysis, 2nd edn. Wiley, Chichester (2002)
2. Royston, P., Altman, D.G., Sauerbrei, W.: Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat. Med.* 25(1), 127–141 (2006)

3. Hornik, K., Stinchcombe, M., White, H.: Universal approximation of an unknown mapping and its derivatives using multilayer neural networks. *Neural Networks* 3, 551–560 (1990)
4. Hastie, T.I., Tibshirani, R.I.: *Generalized Additive models*. CRC Press, Boca Raton (1990)
5. Candiani, G.B., Danesino, V., Gastaldi, A., Parazzini, F., Ferraroni, M.: Reproductive and menstrual factors and risk of peritoneal and ovarian endometriosis. *Fertil. Steril.* 56, 230–234 (1991)
6. Nisolle, M., Donnez, J.: Peritoneal endometriosis, ovarian endometriosis, and adenomyotic nodules of the rectovaginal septum are three different entities. *Fertil. Steril.* 68, 585–596 (1997)
7. Anders, U., Korn, O.: Model selection in neural networks. *Neural Networks* 12, 309–323 (1999)
8. Dehejia, R.: Practical propensity score matching. *J Econometr.* 125(1-2), 355–364 (2005)
9. Hegland, M., McIntosh, I., Turlach, B.A.: A parallel solver for generalised additive models. *Comp. Stat. Data. Anal.* 31, 377–396 (1999)

Recognition of the Ventilatory Response to the Intermittent Chemical Stimuli in Awake Animals

Beata Sokolowska¹, Agnieszka Rekawek¹, and Adam Jozwik²

¹ Medical Research Center, Polish Academy of Sciences, Warsaw, Poland

² Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, Warsaw, Poland
ajozwik@ibib.waw.pl

Summary. In this study we examined whether it could be possible to recognize a type of chemical stimuli, given in intermittent cycles, on the basis of observed changes in the breathing pattern in an animal model. Ventilatory responses to three chemical stimulus - normoxic cycles (3-min administration of stimulus/8-min normoxic recovery) in awake rats were investigated. Two types of chemical stimuli were given: (a) gas mixtures of 14% or 9% O₂ in N₂ (i.e. hypoxia), and (b) 5% or 10% CO₂ in O₂ (i.e. hypercapnia), each one in a separate run of the intermittent stimulus. Ventilatory features: respiratory frequency, tidal volume, minute ventilation, inspired and expired times, were used for recognition of ventilatory responses to exposures of the intermittent stimuli. The quality of recognition was evaluated by a probability of misclassification that was estimated experimentally. As a classifier we used the k nearest neighbor (k -NN) rule that is one of most powerful method offered by the pattern recognition theory. Satisfactory recognition was obtained for recovery periods and stronger stimuli. The best recognition was observed for the intermittent hypercapnia. In conclusion, the approach based on k -NN rule has appeared to be useful tool for recognition changes of ventilatory responses to exposures of the intermittent chemical stimuli.

1 Introduction

Intermittent chemical stimuli, as intermittent hypoxia (IH) or intermittent hypercapnia (IC) are used in studies on humans (healthy subjects or patients), animals and cell cultures [1, 2, 3, 4, 5, 6, 7, 8, 9]. These intermittent stimuli trigger molecular, cellular, and physiological or pathophysiological responses/adaptations, which (a) allow the creating of experimental models of a various diseases that may be observed in human clinic [1, 2, 3, 10, 11, 12] and (b) may be used as a beneficial and a helpful method in training or rehabilitation procedures [13, 14, 15].

In relation to (a), for example, transient episodes of hypoxia are associated with a sleep-disordered breathing manifested as recurrent apneas RA (obstructive or central apneas). A important progress in the field of apnea research was the demonstration that exposing experimental animals to chronic intermittent hypoxia (CIH) elicited the physiological changes similar to that described in recurrent apnea patients [2, 3]. For example, Fletcher and colleagues developed an intermittent hypoxia rat model which mimicking the pattern of hypoxic

episodes encountered during apneas [3]. In the first experiment that used IH, the rats were subjected to intermittent hypoxia as 20s of 5% inspired O₂ for 9 episodes/h 8h/day over 35 days.

It is known, that patients with RA experience not only chronic intermittent hypoxia but also chronic intermittent hypercapnia [16]. Moreover, repeated episodes of intermittent hypercapnia usually accompany the early stages of lung disease. For example, Gozal and workers [17] studied ventilatory responses to six repeated short hypercapnic challenges (5% CO₂ in O₂) in ten healthy adult awake subjects. In their investigations, the breathing variables were measured before, during, and 5 min after administration of the 5% CO₂ for 2 min. Similar research methods are used in experimental models in animals.

In relation to (b), the intermittent hypoxia training (IHT) is often used in sports and medicine (especially in a respiratory rehabilitation) [13, 14, 15]. For example, in physiological studies of acclimatization and of intermittent hypoxic training in human, typical IHT protocols last at least 5 days and consist of approximately 60 min per day, at stimulated altitudes of 3800-5500m. The intermittent hypoxia protocols have a number of potential benefits. It has been shown that in humans the IHT provided increasing of hemoglobin and reducing of blood cell number, and increasing of exercise ventilation and saturation in hypoxia, and also the reducing of the severity of acute mountain sickness.

Other experimental studies have shown that the intermittent hypoxia/ hypercapnia also induced alterations in the respiratory control that reflect various types of neuroplasticity of respiratory output/input [18, 19, 20, 21, 22, 23, 24]. For example, a respiratory long-term facilitation (LTF) is a model of serotonin-dependent plasticity induced by IH. Moreover, neurotrophins such as brain-derived neurotrophic factor (BDNF) and neurotrophin-3 (NT-3) play key roles in many forms of plasticity, also in the respiratory one [1, 4, 21, 22]. Hence, the intermittent chemical stimuli are used in studies of the plasticity phenomenon in experimental models.

The ideal IH or IC research/training protocol or pattern of exposures of intermittent stimuli in animal or human studies is not yet known. In recent years, development of various experimental approaches, with the use of these stimuli provided new interesting insights in to functions of the respiratory system. Acute intermittent hypoxia or hypercapnia, consisting of the hypoxic/hypercapnic episodes frequently occurs in healthy and pathologic situations, but it is rather rarely studied in animals. In the present study we have dealt with recognition of ventilatory responses to acute intermittent hypoxia and intermittent hypercapnia in the awake animal model, by used the k -NN rule.

2 Materials and Methods

Biological Experiments

The study was approved by a local Ethics Committee.

Six experiments were performed on awake adult male Wistar rats, weighing 381 ± 39 g. The animals were placed in plethysmography chamber (Whole Body Plethysmography, WBP System). The WBP system consists of: a chamber of the whole body plethysmograph (Bias Flow Supply PLY3223 with Tether), a preamplifier (the MAXII preamplifier unit), the A/D card, and the BioSystem XA for Windows Software for real time data analysis. This system measures spontaneous ventilation in conscious, unrestrained rats (BUXCO Electronic, Inc., USA). The animals were exposed to intermittent hypoxia (IH) or to intermittent hypercapnia (IC). These chemical stimulus tests were given as gas mixtures of 14% O₂ in N₂ (IH 14%) or 9% O₂ in N₂ (IH 9%) for intermittent hypoxia, or as gas mixtures of 5% CO₂ in O₂ (IC 5%) or 10% CO₂ in O₂ (IC 10%) for intermittent hypercapnia. Five pulmonary variables from a respiratory flow signal of 20 breaths of last minute of exposure and recovery periods were measured: (a) respiratory frequency (f), (b) tidal volume, the inspired volume (TV), (c) minute ventilation (MV, the product of f and TV), (d) inspiratory time (Ti) and (e) expiratory time (Te). The protocol of intermittent stimuli (the intermittent stimulus test) consists of three cycles of 3-min exposure of stimulus (defined as an exposure phase, Exp) and then 8-min normoxic recovery (defined as a recovery phase, Rec).

Data Set

Baseline level (Control) was evaluated before exposures to the intermittent hypoxia and the intermittent hypercapnia. The stimuli (Biological Experiments) were given for 3 min (the exposure phase, Exp) and then followed 8-min normoxic recovery period (the recovery phase, Rec). In summary, each of single intermittent test consisted of seven periods: (a) baseline level (Control), (b) first exposition (Exp1), (c) first recovery (Rec1), (d) second exposition (Exp2), (e) second recovery (Rec2), (f) third exposition (Exp3), (g) third recovery (Rec3). Five ventilatory variables (features: f, TV, MV, Ti, Te) were analyzed in each period. The data set have 140 objects, the 20 objects in each period for each the intermittent stimulus test for each animal, independently.

Analysis Methods

The data analysis was performed with the use of pair-wise classifier based on the k-NN rule and the *leave-one-out* methods [25, 26]. The multi-class pattern recognition task can be decomposed into several binary tasks. So, the multi-decision classifier consists of $nc \cdot (nc-1)/2$ component two-decision classifiers, where nc denotes the number of considered classes. Each of the component classifiers decides between two classes only. The final decision is created by voting of these component classifiers, which are based on k -NN rule with k 's established experimentally by use the *leave-one-out* method. Feature selection was performed separately for each component classifier by reviewing all possible feature

combinations since the primary number of features is small. The selected feature set was obtained by gathering all features selected for the component classifiers.

3 Results and Discussion

Results of the performed analysis are presented in Tables 1, 2, 3. Table 1 consists misclassification errors concerned differentiation between each of exposure or of recovery phases versus baseline level (Control). The error rate of recognition between the periods and Control is given in the last row of Table 1. Error rates of recognition of tree cycles with the use of ventilatory features, taken into account separately for exposure and recovery phases, and jointly for both of them, are presented in Table 2. Misclassification rates of recognition of strength of stimulus, for intermittent hypoxia (IH) and intermittent hypercapnia (IC), are shown in Table 3.

According to the results presented in Table 1, the ventilatory response of presentation time-dependent different stimuli is weaker after hypoxic stimuli than hypercapnic ones. The k -NN classifier allowed to very good differentiation of the ventilatory response of stronger stimuli and also both of intermittent hypercapnia. The error rates shown in Table 3, where stimulus strength was differentiated separately for intermittent hypoxia and intermittent hypercapnia, confirm the results of Table 1. In this case, the k -NN classifier offered small error rates as for features measured in stimulus phase as well in the recovery one. As it was expected, the greater errors concerned the hypoxic stimuli (Table 3). The misclassification rates for the recognition of the cycles of intermittent stimuli obtained in exposure phase, normoxic phase and these phases together, were range from 8.4% to 20.0% - without feature selection, and they were smaller,

Table 1. Misclassification rates (E_r). Recognition of ventilatory response between exposure (Exp_i) or recovery (Rec_i) ($i=1,2,3$, cycle number) and control level (i.e. before application of intermittent stimuli, Control) for intermittent hypoxia (of IH 14% and 9%) and intermittent hypercapnia (of IC 5% and 10%) without and with feature selection (in parenthesis).

Phases of stimulus & cycles (no 1,2,3)	Intermittent hypoxia Vs. Control [E_r]		Intermittent hypercapnia Vs. Control [E_r]	
	IH 14%	IH 9%	IC 5%	IC 10%
Exp ₁	0.138	0.000	0.000	0.000
Rec ₁	0.088	0.088	0.063	0.050
Exp ₂	0.163	0.000	0.000	0.000
Rec ₂	0.113	0.000	0.038	0.000
Exp ₃	0.050	0.000	0.000	0.000
Rec ₃	0.013	0.013	0.063	0.000
Summary recognition (<i>after feature selection</i>)	All phases: 0.268 (<i>0.193</i>)	All phases: 0.158 (<i>0.107</i>)	All phases: 0.150 (<i>0.097</i>)	All phases: 0.115 (<i>0.063</i>)

Table 2. Misclassification rates (E_r). Recognition of cycles, in exposure (Exp), recovery (Rec) and the both phases, for intermittent hypoxia (of IH 14% and 9%) and intermittent hypercapnia (of IC 5% and 10%), without and with feature selection.

Phases of intermittent stimulus	Intermittent hypoxia Recognition of 3 cycles [E_r]		Intermittent hypercapnia Recognition of 3 cycles [E_r]	
	IH 14%	IH 9%	IC 5%	IC 10%
<i>Without selection:</i>				
Exp	0.200	0.217	0.159	0.134
Rec	0.150	0.084	0.117	0.100
Exp&Rec	0.208	0.150	0.138	0.117
<i>With selection:</i>				
Exp	0.117	0.175	0.117	0.075
Rec	0.117	0.034	0.067	0.067
Exp&Rec	0.142	0.104	0.092	0.071

Table 3. Misclassification rates (E_r). Recognition of stimulus strength, in exposure (Exp) and recovery (Rec) phases, for intermittent hypoxia (between IH 14% and 9%) and intermittent hypercapnia (between IC 5% and 10%), without and with feature selection.

Phases of intermittent stimulus	Intermittent hypoxia Recognition of stimulus strength (14% vs. 9%) [E_r]	Intermittent hypercapnia Recognition of stimulus [E_r] strength (5% vs. 10%) [E_r]
	<i>Without selection:</i>	
Exp	0.042	0.000
Rec	0.180	0.083
<i>With selection:</i>		
Exp	0.029	0.000
Rec	0.029	0.067

from 3.4% to 17.5% - after feature selection (Table 2). The recognition of cycles was better in the recovery phase than in the exposition one. In summary, good recognition of cycles in the recovery phases suggests that after exposure of the intermittent stimuli the breathing pattern is not as control one and furthermore the ventilatory response changes in the sequential cycles.

The paper demonstrated preliminary results concerned to the study of the ventilatory responses to short-time intermittent stimuli in the awake rats. The results confirm that the use of pattern recognition methods can be a very useful tool in studies of early respiratory changes after application the intermittent stimuli. In fact, the authors intend to carry out further researches on neuroplasticity changes in the respiratory system, after the acute/chronic or short/long-time exposures of IH or IC in the anesthetize/awake animal models.

Acknowledgement. This work was supported by the statutory budget of the Medical Research Center and the Institute of Biocybernetics and Biomedical Engineering of the Polish Academy of Sciences.

References

1. Prabhakar, N.R., Fields, R.D.: Intermittent hypoxia: cell to system. *Am. J. Physiol. Lung. Cell Mol. Physiol.* 281, 524–528 (2001)
2. Neubauer, J.A.: Invited review: physiological and pathophysiological responses to intermittent hypoxia. *J. Appl. Physiol.* 90, 1593–1599 (2001)
3. Fletcher, E.C.: Invited review: Physiological consequences of intermittent hypoxia: systemic blood pressure. *J. Appl. Physiol.* 90, 1600–1605 (2001)
4. Kinkead, R., Bach, K.B., Johnson, S.M., Hodgeman, B.A., Mitchell, G.S.: Plasticity in respiratory motor control: intermittent hypoxia and hypercapnia activate opposing serotonergic and noradrenergic modulatory systems. *Com. Biochem. Physiol.* 130, 207–218 (2001)
5. Baumgardner, J.E., Otto, C.M.: In vitro intermittent hypoxia: challenges for creating hypoxia in cell culture. *Respir. Physiol. Neurobiol.* 136, 131–139 (2003)
6. Prabhakar, N.R., Kumar, G.K.: Oxidative stress in the systemic and cellular responses to intermittent hypoxia. *Biol. Chem.* 385, 217–221 (2004)
7. Reeves, S.R., Gozal, D.: Changes in ventilatory adaptations associated with long-term intermittent hypoxia across the age spectrum in the rat. *Respir. Physiol. Neurobiol.* 150, 135–143 (2006)
8. Sokółowska, B., Pokorski, M.: Ventilatory augmentation by acute intermittent hypoxia in the rabbit. *J. Physiol. Pharmacol.* 57(4), 341–347 (2006)
9. Mahamed, S., Mitchell, G.S.: Is there a link between intermittent hypoxia-induced respiratory plasticity and obstructive sleep apnoea? *Exp. Physiol.* 92(1), 27–37 (2007)
10. Sica, A.L., Greenberg, H.E., Ruggiero, D.A., Scharf, S.M.: Chronic-intermittent hypoxia: a model of sympathetic activation in the rat (2000)
11. Campen, M.J., Shimoda, L.A., O'Donnell: Acute and chronic cardiovascular effects of intermittent hypoxia in C57BL/6J mice. *J. Appl. Physiol.* 90, 1593–1599 (2005)
12. Zieliński, J.: Effects of intermittent hypoxia on pulmonary haemodynamics: animal models versus studies in humans. *Eur. Respir. J.* 25, 173–180 (2005)
13. Levine, B.D.: Intermittent Hypoxic Training: Fact and Fancy. *High Altitude Med. Biol.* 3(2), 177–193 (2002)
14. Serebrovskaya, T.V.: Intermittent hypoxia research in the former Soviet Union and the commonwealth of independent states: history and review of the concept and selected applications. *High Altitude Med. Biol.* 3(2), 205–221 (2002)
15. Serebrovskaya TV, Swanson RJ, Kolesnikova EE (2003) Intermittent hypoxia: mechanisms of action and some applications to bronchial asthma treatment. *J Physiol Pharmacol* 54(Supp.I): 35-41
16. Rapoport, D.M., Garay, S.M., Epstein, H., Goldring, R.M.: Hypercapnia in the obstructive sleep apnea syndrome. A reevaluation of the "Pickwickian syndrome". *Chest* 89(5), 627–635 (1986)
17. Gozal, D., Ben-Ari, J.H., Harper, R.M., Keens, T.G.: Ventilatory responses to repeated short hypercapnic challenges. *J. Appl. Physiol.* 78(4), 1374–1381 (1995)
18. Mitchell, G.S., Backer, T.L., Nanda, S.A., Fuller, D.D., Zabka, A.G., Hodgeman, B.A., Bavis, R.W., Mack, K.J., Olson, E.B.: Invited review: intermittent hypoxia and respiratory plasticity. *J Appl. Physiol.* 90, 2466–2475 (2001)

19. Gozal, E., Gozal, D.: Invited review: respiratory plasticity following intermittent hypoxia: developmental interactions. *J. Appl. Physiol.* 90, 1995–1999 (2001)
20. Ling, L., Fuller, D.D., Bach, K.B., Kinkead, R., Olson, E.B., Mitchell, G.S.: Chronic intermittent hypoxia elicits serotonin-dependent plasticity in the central neural control of breathing. *J. Neuroscience* 21(14), 5381–5388 (2001)
21. Backer-Herman, T.L., Mitchel, G.S.: Phrenic long-term facilitation requires spinal serotonin receptor activation and protein synthesis. *J. Neuroscience* 22(14), 6239–6246 (2002)
22. Backer-Herman, T.L., Fuller, D.D., Bavis, R.W., Zabka, A.G., Golder, F.J., Doperalski, N.J., Johnson, R.A., Watters, J.J., Mitchell, G.S.: BDNF is necessary and sufficient for spinal respiratory plasticity following intermittent hypoxia. *Nature Neuroscience* 7(1), 48–55 (2004)
23. Reeves, S.R., Gozal, D.: Developmental plasticity of respiratory control following intermittent hypoxia. *Respir. Physiol. Neurobiol.* 149, 301–311 (2005)
24. Dahan, A., Nieuwenhuijs, Teppema, L.: Plasticity of central chemoreceptors: effect of bilateral carotid body resection of central CO₂ sensitivity. *Acta Physiol. Sinica.* 59(2), 128–132 (2007)
25. Devijver, P.A., Kittler, J.: *Pattern Recognition: A Statistical Approach*. Prentice Hall, London (1982)
26. Duda, R.O., Hart, P.E., Stock, D.G.: *Pattern Classification*. John Wiley and Sons, New York (2001)

Telesfor – Telemedical Real-Time Communication Support System

Jerzy Błaszczczyński, Bartłomiej Prędko, and Roman Słowiński

Institute of Computing Science, Poznań University of Technology,
60-965 Poznań, Poland

{jblaszczynski, rslowinski, mszelag}@cs.put.poznan.pl

Summary. Telesfor system is the result of an initiative undertaken by researchers from Microsoft Innovation Center in Poznan, that attempts to bring a user-friendly and safe system for the support of communication in healthcare. Telesfor is based on ConferenceXP architecture. It enables a document-based collaboration that is enhanced by exchange of text messages and videoconferences between multiple participants of a communication session. It also brings tools for effective organization of a session allowing to address differences in role and competence of participants. In this paper, we present Telesfor system and show how it may be used in telemedical consultations and in distance learning.

1 Introduction

In the course of development of telemedical portal "Telemedycyna Wielkopolska" [4, 5], in the fall of 2006, we faced the need of a safe tool that would allow interactive collaboration between physicians on medical documents. This collaboration was to be enriched by exchange of text messages and videoconferencing. The subject for the collaboration were mainly medical images (i.e., digital X-rays, US scans, CT (computer tomography), MRI (magnetic resonance imaging), fMRI (functional MRI) and other image clinical documentation) coming from examination of medical cases. An expected feature of this tool was its ease of use and ability to run it independently from the portal framework. Also the documents were expected to come from different sources with main requirement being the ability to share them easily between participants. The documents were to be consulted by a group of specialists that were located in different departments of a hospital. They could have also been a base for a specialized remote consultation of difficult medical cases encountered in regional hospitals that require to be consulted remotely with experts in a specialized clinic. Medical tele-education and collaborative learning were another applications of the tool expected by physicians. This function was supposed to allow students and medical personnel to constantly widen their knowledge and thus improve the overall quality of the health care. The result of our development is the system for the support of communication in health care called Telesfor.

The structure of this document is the following. In section 2, we present the motivations and the expected functionality of the Telesfor system. In section 3, we present technologies in use. In section 4, we show Telesfor's architecture. In section 5, we describe in depth the current functionality of the system and we show an example of its use. We conclude our presentation in section 6.

2 Motivation and Expected Functionality

Today, we are seeing the emergence of devices and technology that support collaborative communication in ways we have envisioned for many decades. With wireless and high-bandwidth networks, enhanced with high-quality, low-latency audio and video, participants of this communication processes can collaborate in interactive workspaces, both synchronously and asynchronously. Additionally, these networks enable an easy access to content, collaborators, experts, mentors, and laboratories, so participants can truly work from anywhere.

Our approach to develop Telesfor system was inspired by the research on Microsoft's advanced collaboration architecture ConferenceXP [3], Classroom Presenter system developed at the University of Washington [1, 2] and by our experience with clinical decision support systems [6]. Telesfor was expected to fulfil the following requirements, which meet both expectations of physicians and possibilities of the available technologies:

- synchronous and asynchronous collaboration of multiple participants,
- teacher and student modes of participation,
- videoconferencing and exchange of text messages,
- image, video and signal documents sharing,
- image processing and image enhancement procedures (i.e., histogram equalization, convolution filtering, noise suppression),
- measurements of image features and objects (e.g., span, area),
- image annotation and signalization features,
- secured and archived collaboration stored and managed by session mechanism

3 Technologies

The aforementioned requirements were considered in view of the available technologies. We looked for a technology that would provide us with good communication architecture and that would be flexible enough to adapt and extend it to our specific needs. First we have checked the possibility to extend the functionality of the typical communication programs like MSN Messenger or Skype. They are mostly designed for point-to-point communication and programming their extensions is hardly achievable. We have also checked open-source protocols, namely Jabber. This, although promising, lacked required sophistication. Mainly the multi-user voice conference was causing problems. Another possibility was to use Microsoft Live Communications Server. It is a server based application

for voice communication implementing Simple Internet Phone protocol, which also lacked more advanced collaboration functions.

Finally, we focused our interest on a Microsoft Research project, called ConferenceXP, that is designed as an advanced collaboration and distance learning platform for schools and universities. It allows the parallel collaboration of many participants on a series of documents (preferably images and presentation slides) with accompanying voice and video communication. We have chosen this technology because it satisfied our needs and it is easily expandable. Moreover, ConferenceXP’s source code is publicly available, making it a good base for an academic research project.

3.1 ConferenceXP

ConferenceXP concentrates on features called venues. Venue is a virtual place in a network where people meet to collaborate on a subject. Venue service shown in Fig. 1, is responsible for management of venues. Any participant in the venue has access to equal functionality. Anyone can add images, speak or send a video stream.

ConferenceXP requires the network it operates on to support multicast architecture. Some of the participants can also connect from the unicast network by use of a reflector service. Multicast architecture reduces the volume of data transferred between client computers. ConferenceXP basically adapts a peer-to-peer communication scheme. Venue service is used only for discovery of venues to which a client can connect. All communication is performed between clients and a multicast address. Specialized archive service is used to store all data sent during the venue session in the database server, including voice and video stream.

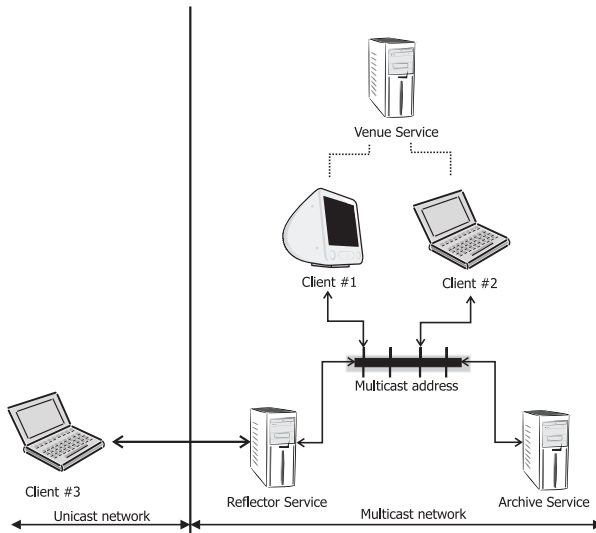


Fig. 1. ConferenceXP architecture

The ConferenceXP Network Transport layer provides custom-written network transport technology that ensures audio, video, and data streams transmission with a minimum loss of data. ConferenceXP sends data streams over the network by using an implementation of the Real-time Transport Protocol (RTP), which is an IETF standard.

ConferenceXP is written for Microsoft .Net Framework 2.0 in C# language. Its functions are grouped in so called capabilities, for example text message exchange capability or presentation capability. Software developers can create new capabilities using available ConferenceXP application programming interface (API).

3.2 Other Technologies

ConferenceXP uses Microsoft Ink technology, designed for tablet computers, for making image annotation. These, called strokes, can be made using the mouse or the pen on the TabletPC platform. Strokes are sent to all participants and stored in the archive server. Microsoft Ink technology works also on regular Windows systems. Ability to interoperate with Tablet PC platform is an important feature since this allows to incorporate active learning techniques into ConferenceXP.

We have decided to use Microsoft SQL Server 2005 as the database back-end for Telesfor system. Database engine is used for storing session information and session archiving.

4 Telesfor's Architecture

ConferenceXP seemed to be the perfect founding stone for a medical communication support system. However, we have realized soon that it lacks several important features that are required for medical applications. For example, venue is not recording any information about the state of an ongoing session. Thus, a newly connected participant is provided with no information on what happened before (s)he joined in. If any of participants loses the connection to the session and connects again, all information shared in between is lost for him. ConferenceXP does not provide user authentication services. Anyone can connect to a session. Moreover, data transmitted during the session is not protected in any way. Lack of each of these features is not acceptable in medical environment.

We decided to build upon the ConferenceXP and to provide all of the required features. Thus, first we have modified the ConferenceXP architecture. The modifications, shown in Fig. 2, include new elements in Telesfor's architecture:

- consultation server – responsible for creation and management of teleconsultation sessions,
- database server – used to store all data about sessions,
- monitor client – specialized type of client used to monitor the session lifetime and store its state in the database.

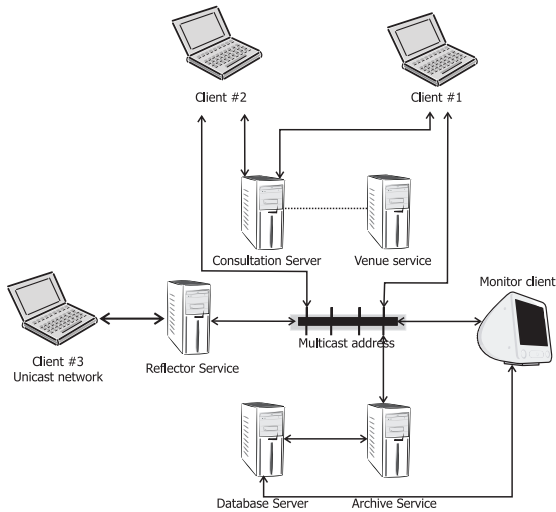


Fig. 2. Telesfor architecture

All software elements mentioned above can run on a single computer, preferably on Microsoft Windows 2003 Server operating system. We have also designed from scratch a new client application, called Telesfor client, that is Microsoft Windows XP and Microsoft Vista compatible.

According to requirements of the Telesfor system, different modes of users' participation in the session are distinguished. We decided to divide the participants into three groups: creator/owner of the session, lecturer, and regular participant.

Creator of the session has all the privileges, whereas regular participant can only comment on the session subject. Lecturer can turn on the session share mode. In this mode every participant sees the same content at a time. Lecturer privilege can be granted to any of the participants in the session. Additionally, a participant appointed as the lecturer can use a virtual laser pointer to draw attention of other participants to the interesting area of the document. (s)he can also make an annotation to the image if (s)he wants the area to be permanently marked.

One of our goals was to provide a comprehensive set of tools for image manipulation. Telesfor includes a typical set of image enhancement functions like brightness and contrast adjustment, changes to the color depth or a set of linear filters. The results of application of image manipulation tools can be shared between participants of sessions.

Telesfor can work with images stored in most of popular formats, as well as with DICOM images. DICOM is the medical standard for image storing and retrieval. Telesfor image viewer can open DICOM images and is able to perform length and area measurements, assuming that spatial resolution information is

embedded into the image. All DICOM images are stored and distributed between session participants in their original form.

5 Session Example

A session during which DICOM images are consulted between participants illustrates a typical use of the Telesfor system. The session begins when participants join the venue and upload the documents that they intend to share. As it was mentioned, participants are not required to come to the session at the same time. Telesfor takes care, so that latecomers are provided with the results of ongoing session. Also not all documents have to be shared from the beginning of the session. Once those initial stages are finished, all of the participants can see the subjects of their collaboration, the list of shared images on the left of the main window of Telesfor, as it is shown in Fig. 3. On the right of the main window, a list of participants and a window with text messages are displayed.

Usually, participants exchange text messages with some first indications of their intentions at the beginning of the session. At any time during the session, they may also start a videoconference. Each of the participants may choose to send and/or receive audio and visual content. Fig. 3, shows session screen during a videoconference.

As the session continues, participants begin to show important parts of discussed documents by annotations in images. This is an efficient mechanism allowing to focus attention of participants on selected parts of the images. An example of annotations in a US scan is shown in Fig. 3. Participants may also accurately measure selected objects in discussed images. Measurements are done by ‘pointing and clicking’ or by using area selection. This feature is, however, applicable to DICOM images only. Other types of images do not provide information about

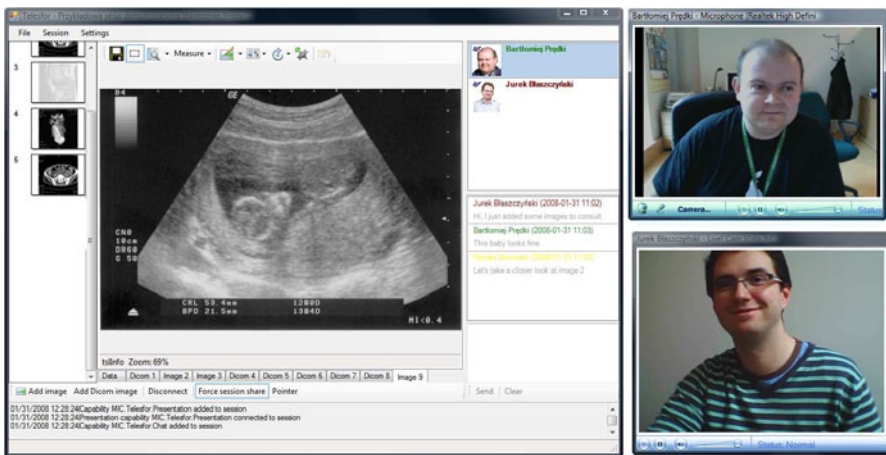


Fig. 3. Telesfor – consultation session screen

physical size of objects that they contain. Results of the measurement process are only seen by those of participants that wish to participate in it.

Any of participants can choose to operate in shared mode where one of participants appointed as the lecturer is guiding others and presenting his/her opinions. Every participant being in this mode sees the same image and the same operations on his/her screen. This feature is particularly useful in distance learning sessions. In such a case, participants being students can follow the presentation that is given by the lecturer. Naturally, some of participants may alternatively focus on other aspects of consultation. For example, one of participants may want to annotate and possibly make some measurement on his/her own. (s)he can disable shared mode and work separately. Then, at any time (s)he can come back to shared mode to present his/her results to others. (s)he can also be appointed as the lecturer when wishes to. Obviously, usage of shared mode and lecturer mode requires some discipline from participants.

Image enhancement procedures are mandatory when analyzing medical images. Each of participants may apply those procedures according to his/her own needs. Application of those procedures can also be made in shared mode. Image enhancement procedures include simple methods (e.g., automatic contrast adjustment) and more advanced methods (e.g. noise suppression, convolution filtering, pseudocoloring). Participants apply those procedures to reveal details that were hidden in the original images. This can lead to some new observations that need to be shared further. It is participants' choice when to finish the session.

Satisfied with the results of the consultation, participants can move to another case and start a new consultation session. The session is archived in the database. Participants can come back to it or even continue the session at any time later.

6 Summary

We have described the results of the work on Telesfor system. We have also presented an example of a typical session that shows the features of the system and their usefulness during consultation and distance learning. Telesfor was used during test consultation sessions by physicians from Poznań University of Medical Sciences. It is also planned to be used in an educational setting with students from this university.

We are looking forward to enhancing Telesfor with processing of signal data mainly in the form of ECG (electrocardiogram) documents. More advanced sharing of multimedia documents will be also enabled. This should allow sharing video documents with a possibility to synchronise their presentation during a session. Possibility to store some parts of a session as Microsoft PowerPoint slides is also planned. This should allow for a consultation session to become a valuable part of presentation slides shown during a lecture.

Acknowledgement. The authors acknowledge financial support from Microsoft Innovation Center in Poznań.

References

1. Anderson, R., Anderson, R., McDowell, L., Simon, B.: Use of Classroom Presenter in Engineering Courses, *Frontiers in Education* (2005)
2. Anderson, R., Anderson, R., Davis, P., Linnell, N., Prince, C., Razmov, V., Videon, F.: Classroom Presenter: Enhancing Interactive Education with Digital Ink. *IEEE Computer* 40, 56–61 (2007)
3. Beavers, J., Chou, T., Hinrichs, R., Moffatt, C., Pahud, M., Powers, L., Van Eaton, J.: The Learning Experience Project: Enabling Collaborative Learning with ConferenceXP, Microsoft Research Technical Report, MSR-TR-2004-42 (2004)
4. Błaszczczyński, J., Kosiedowski, M., Mazurek, C., Słowiński, R., Słowiński, K., Stroiński, M., Wilk, S.z.: Telemedical portal 'Telemedycyna Wielkopolska'. In: Piętka, E., Leski, J., Franiel, S. (eds.) *Medical Informatics & Technology*, pp. 230–235. MIT Press, Cambridge (2006)
5. Błaszczczyński, J., Kosiedowski, M., Mazurek, C., Wilk, S.z.: Ontologies for knowledge modeling and creating user interface in the framework of telemedical portal. In: Stormer, H., Meier, A., Schumacher, M. (eds.) *European Conference on eHealth 2006. Proceedings of the ECEH 2006, Lecture Notes in Informatics, Gesellschaft fur Informatik*, vol. P-91, pp. 275–286 (2006)
6. Michałowski, W., Słowiński, R., Wilk, S.z., Farion, K., Pike, J., Rubin, S.: Design and development of a mobile system for supporting emergency triage. *Methods of Information in Medicine* 44, 14–24 (2005)

Multimedia Program for Teaching Autistic Children

Joanna Marnik¹ and Magdalena Szela^{2,*}

¹ Rzeszow University of Technology, ul. W. Pola 2, 35–959 Rzeszów
jmarnik@prz-rzeszow.pl

² Special School, ul. Marszałkowska 24s, 35–215 Rzeszów
mbszelka@wp.pl

Summary. Autism is an incurable neurological disease. The main symptoms of this disease are problems with communication and social behaviour. There are difficulties with integration of sensory impressions, too [1]. Nowadays, the only way of solving the problem of autism is rehabilitation, the aim of which is for an autistic person to achieve the best level of functioning in a society.

On account of communication problems of autistic people in their education, it is essential to use methods which allow a teacher to move away to the further plan. This is possible by means of virtual reality methods. This kind of computer program is presented in this paper. The program is designed for rehabilitation of autistic children. The main aim of the program is to show typical human behaviour in daily situations and to make the autistic children familiar with emotions expressed by facial expressions.

1 Introduction

Autism is a comprehensive development disorder. This illness has the influence on all spheres of child's functioning. The typical irregularities in autism are:

- language development and communication disorders – the inability to initiate and continue conversation, and the inability to read nonverbal messages,
- disturbances in contacts with others and in social interactions – children's behaviour is schematic and not in accordance with widely accepted social norms,
- functional and symbolic games.

It is estimated that 4–5 in 1000 people suffer from autism, mainly boys [1]. They do not have the sufficient knowledge about how they should behave in specified situation. They have a problem with understanding other people's intentions and emotions. A visual and auditory memory is usually the strong point in

* We are grateful to Lukasz Warda for his valuable help in preparing the application.

the functioning of autistic people. Therefore, visualisation of actions plays an essential role in their therapy [2].

Neglecting the problems at the early stage usually leads to the child's retardation. Therefore, diagnosing the illness at the early stage of childhood and introducing rehabilitation programs as fast as possible is very important. Multimedia programs and virtual reality are becoming more and more popular in rehabilitation of children with different kinds of developmental disorders. One reason for this is that a person who carries out the rehabilitation can be replaced by virtual objects and characters. Furthermore, the attractiveness of learning and, by implication, motivation of a child to work on their problems, is greater when an interactive computer program is used. Additionally, it is possible to adjust the system of awards to an individual child. It is especially important when working with children who have an inclination to passive behaviour, like in autism.

Multimedia and virtual reality have been used for some time now in diagnosing, teaching and rehabilitation of children with different kinds of developmental problems. Lányi and Tilinger [3] have developed a multimedia and virtual reality software package for the rehabilitation of autistic children. The package contains a virtual environment helping to learn how to do the shopping, and two kinds of interactive multimedia software for teaching how to get dressed and use public transport. Noris and others [4] proposed a computer based approach to the analysis of social interaction environments for diagnosis of Autism Spectrum Disorders (ASD) in young children of 6–18 months of age. They applied face detection on videos from a head-mounted wireless camera to measure the time a child spends looking at people. Other system [5] was designed to stimulate the social attention of pre-school-aged children with ASD. The system uses eye-tracking to determine the object on which the examined child focuses their attention. The child is sitting in a children's arcade helicopter equipped with a flat screen monitor and an eye-tracking camera. The objective of the training is to draw the child's attention to a face displayed on the screen. Looking at the face, extracting information from it and other nonverbal behaviour, and reacting to this information are rewarded by showing to the child their favourite videos.

The multimedia computer program has been designed to teach autistic children typical patterns of human behaviour during everyday tasks. The main aim of the program is to improve communication skills of the children. It is accomplished by demonstrating typical behaviour of a person who gets in touch with another person, and how he or she continues the conversation. The program allows to acquaint the child with facial gestures and to show what successive actions must be taken to e.g. rent a film from a video shop. Additionally, corresponding pictograms [6] are introduced. The program's graphic side was built on the basis of the graphic engine of the well-known computer game *The Sims 2* [7]. The application is presented in section 2. It is used in classes for autistic children of the Special School in Rzeszow. Its effectiveness assessed on the basis of the work with autistic pupils is shown in section 3. At the end of the paper the conclusion of the work is presented (Sec. 4).

2 The Application

The main objective of the following application is to enhance communication skills of children with autism. It is done by demonstrating to them ways of getting in touch with other people and by acquainting them with facial gestures. Another objective is to teach a child completing chosen tasks, related to their functioning in everyday life, for example renting a film from a video shop. Furthermore, the program helps to acquaint the child with concepts concerning everyday activities and behaviour. Pictograms related to the concepts are also introduced. They are widely used as an alternative or augmentative communication system [6].

The program was created in Microsoft Visual C++ 6.0 environment. Video clips used in it were prepared by means of the graphic engine of well-known computer game *The Sims 2* [7]. The video clips were additionally processed by VirtualDub application [8]. Animation and sound effects are used in the application, too. A lector's voice informs a child about a task, which is to be done, and gives a hint on how to do that. A character who appears on the screen uses universal body gestures and their voice. The presented scenes have a simple structure. The animation fills the entire screen. On the bottom of the screen there are buttons for the child to influence the course of the action. These buttons are simple and clear. They present a pictogram or a short description of their function, depending on the settings. The teacher who monitors the child's work with the program can set up a number of options, thus adjusting its action to the child's abilities and needs. It is possible to arrange the buttons in two ways. The order of the buttons can be random or can correspond to the actions which should be done during the interaction with the program. The buttons can be visible constantly, or they can appear when they are needed and disappear when they are not. The speed of playing the scene can also be adjusted.

2.1 Teaching How to Get in Touch

Autistic children have enormous problems with getting in touch with others, even with the closest family members (parents, siblings). The inability to establish these relations properly, according to widely accepted norms, is the essence of this disorder [9]. Therefore, four scenes teaching a child to react properly to certain situations, ie. in the culturally accepted way, were included in the application.

The scenes differ from one another as for the sex of the people who have an opportunity to get in touch and the complexity of surrounding, in which the scene is supposed to take place. Owing to that the child learns which elements are essential to interact with other people and which are independent of other factors. During the interaction with a given scene the child can affect whether and when characters appearing on the screen get in touch with one another. The child can also decide how the conversation develops. Possible actions are: calling a person, approaching them, introducing oneself, waiting for the response and shaking hands. Each of these actions is initiated by pushing the appropriate button. An example scene, in which two girls are ready to get in touch, is shown in the Fig. 1.



Fig. 1. A scene in a room presenting two girls ready to get in touch

2.2 Acquainting with Facial Gestures

Autistic children do not use mimic possibilities for the nonverbal communication. Moreover, they avoid looking at others. They use gesticulation only occasionally and only in order to achieve their own purpose. These children have also difficulty in interpreting the meaning of gestures, particularly facial gestures. The majority of these children do not express their emotions in any way. The next two of the presented scenes aim at attracting the child's attention to the face of a virtual character shown on the screen. In the first of these scenes the character is male, in the second female. With the help of the facial gestures, this character expresses emotions such as anger, fury, sadness, satisfaction. Recognising the presented emotion is the child's task. Additionally, the pictograms [6] related to the emotions are introduced here. Thanks to these scenes, the child learns to recognise emotions. Fig. 2 presents a virtual girl expressing anger.

2.3 Renting a Film

The main aim of the rehabilitation of the children with developmental disorders is for the children to gain as much independence as possible, so that they can function in a society as well as possible. Teaching these children individual abilities is not enough if the child is not able to apply them practically. At this stage of learning direct instructions are needed. These instructions should give the child a hint on how to connect appropriate actions to achieve the intended goal. Usually the children start to use sequences of particular actions independently



Fig. 2. A scene depicting anger which is expressed by a virtual girl

when they are able to do the task without the teacher's help and when they properly interpret stimuli from the surrounding [9].

The next two scenes were prepared to teach the child what actions, and in what order, should be performed to rent a film from a video shop. These scenes differ only in a virtual character. In one of them it is a girl, in the other it is a boy. The following activities were distinguished here: coming up to the shelf with films, choosing a film, coming up to a cash desk, asking about the price, waiting for a receipt, paying. An example picture from the scene in which a boy rents a film is shown in the Fig. 3.

2.4 Everyday Behaviour and Activities

The autistic children do not often understand the purpose and meaning of the language. Therefore, they are not able to use it as a communication tool. Only a few of them are able to use a language in the right way to express and communicate their thoughts and emotions to others [9]. Hence, it is necessary to introduce alternative means of communication while working with such children. The pictograms [6] are often used here. A given word or concept is replaced by its visual representation. The following three scenes are used to introduce pictograms presenting everyday activities, such as washing up, cleaning, eating, reading, sitting, cooking, learning, etc., and emotions, e.g. cry, fear, acceptance, lack of acceptance, lack of understanding. The child is to show a pictogram which corresponds to what is happening on the screen. An exemplary scene, which presents the action of drinking, is shown in the Fig. 4.



Fig. 3. A scene in which a boy rents a film



Fig. 4. A scene presenting the action of drinking

3 Usability of the Application

The application was prepared in cooperation with teachers of the autistic children. Currently, it is being used in classes for autistic children of the Special School in Rzeszów. Useful tips for further work are provided from the observations of the

children working with the program and discussions with their teachers. Owing to that we hope that next applications will be better adapted to the needs of the autistic children.

In the assessment of the program, the following aspects were taken into consideration [10]:

- user-friendliness,
- motivation of the child to work with the program,
- knowledge acquisition,
- an ability to use the acquired skills in everyday life.

For the child to use the program it is necessary to understand the task and know how to navigate the program by means of the buttons. The task which the child is supposed to do is explained by a lector who replaces the teacher. The program's buttons are operated by indicating them with a cursor and pushing the left button of the mouse. An interaction with the program has been maximally simplified so that the task could be done without the help of the teacher who supervises the child.

As a result, motivating the child to use the program was fairly easy. The child's interest appeared right after starting the game. Examining the effectiveness of the knowledge acquisition by using the presented application and the ability to use the acquired skills in everyday life requires more time. Now some research on these aspects is underway and that is why there are no concrete results yet. However, it is already possible to say that the application has received favourable opinions of both the teachers and the parents of the autistic children. They are sure that this form of teaching will bring significant benefits, because it has been shown that children can transfer knowledge and skills gained in a virtual environment to the real world [11].

4 Conclusions

The autistic children have difficulties in learning mainly due to their inability to get in touch with others. In a further perspective it usually leads to delaying the process of development. The multimedia program presented in the paper is designed for the rehabilitation of autistic children. Its main goal is to develop communication skills of these children.

The experiences gained to this moment show the direction for further work in this field. We are going to develop next applications which will be better adjusted to the needs and skills of the autistic children. The emphasis will be put on a more natural way of interaction with a computer and bigger flexibility in the choice of virtual characters and scenes.

We are sure that the rehabilitation with the use of modern techniques and tools, including virtual reality and multimedia, introduced at an early stage of the development of the child with various kinds of developmental disorders, will provide them with better chances of development.

References

1. Galka, U., Peczkowska, E.: Dzieci z autyzmem. Centrum Metodyczne Pomocy Psychologiczno-Pedagogicznej, Warszawa (2007)
2. Waclaw, W., Aldenrud, U., Ilstedt, S.: Dzieci z autyzmem i zespołem Aspergera. Praktyczne doświadczenia z codziennej pracy, "Ślask", Katowice (2000)
3. Lányi, C.S., Tilinger, Á.: Multimedia and Virtual Reality in the Rehabilitation of Autistic Children. In: Miesenberger, K., Klaus, J., Zagler, W., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 22–28. Springer, Heidelberg (2004)
4. Noris, B., Benmachiche, K., Meynet, J., Thiran, J.-P., Billard, A.G.: Analysis of Head-Mounted Wireless Camera Videos for Early Diagnosis of Autism. In: Kurzyński, M., et al. (eds.) Computer Recognition Systems 2. ACS, vol. 45, pp. 663–670 (2007)
5. Trepagnier, C.Y., Sebrechts, M.M., Finkelmeyer, A., Woodford, J., Steward Jr., W.: Virtual social environment for preschoolers with autism – preliminary data. In: Proc. 6th Intl. Conf. Disability, Virtual reality & Assoc. Tech., Esbjerg, Denmark, pp. 43–49 (2006)
6. Podeszewska-Mateńko, M.: Piktogramy – szwedzki system komunikacji znakowo-obrazkowej, <http://www.plock.edu.pl/prv/logopeda/komunikacja/piktogram1/piktogram1.html>
7. The Sims 2, www.simsy.pl
8. Lee, A.: VirtualDub 1.6.17, <http://www.virtualdub.org>
9. Randall, P., Parker, J.: Autyzm. Jak pomóc rodzinie. Gdańskie Wydawnictwo Psychologiczne (2001)
10. Pantelidis, V.S.: Reasons to Use Virtual Reality in Education, <http://vr.coe.ecu.edu/reas.html>
11. McComas, J., Pivik, J., Laffamme, M.: Current Uses of Virtual Reality for Children with Disabilities. In: Riva, G., et al. (eds.) Virtual Environments in Clinical Psychology and Neuroscience. Ios Press, Amsterdam (1998)

Multimedia System for Accessible Distant Education

Dominik Spinczyk¹ and Piotr Brzoza²

¹ Silesian University of Technology, Department of Biomedical Engineering, Gliwice, Poland

dspinczyk@polsl.pl

² Silesian University of Technology, Department of Informatics, Gliwice, Poland

piotrbr@polsl.pl

Summary. Paper describes a computer system enabling interactive online presentation of multimedia Daisy books over the Internet. The system cooperates with the Internet multimedia library computer management system. The main goal of both projects and their execution, is easy and effective access to information for visually impaired people. We focus on new feature of our DaisyReader which allows interactive voice reading of math formulas.

1 Introduction

Visually impaired people have limited access to information presented in traditional form. Computers with assistive software like as screen readers, screen magnifiers and multimedia browsers enable impaired users potential unlimited access to information. However, information accessibility depends on agronomy of browser's user interfaces and accessibility of digital content.

Internet library portals, friendly to visually impaired readers allow for easy and effective access to catalogues of public and academic libraries. Nowadays more and more libraries offer online access to digital content: eBooks, eMagazines, music and films. Publishers offer digital materials in several different formats: .doc, .xml, .pdf, .lid. To read them users need various e-Book reading software [1].

2 Daisy Standard

We present new multimedia eBook format (DAISY 3.0 ANSI/NISO Z39.86-2005) developed by Daisy Consortium www.daisy.org. Daisy books present book content in multimode form including text, audio, and graphics. Readers can easily navigate in logical book structure by: chapters, headings, pages, paragraphs and sentences. Main daisy book structure is presented in Fig. 1. Daisy books can be played using hardware or software players. Currently world wide there are about 130000 available book titles.

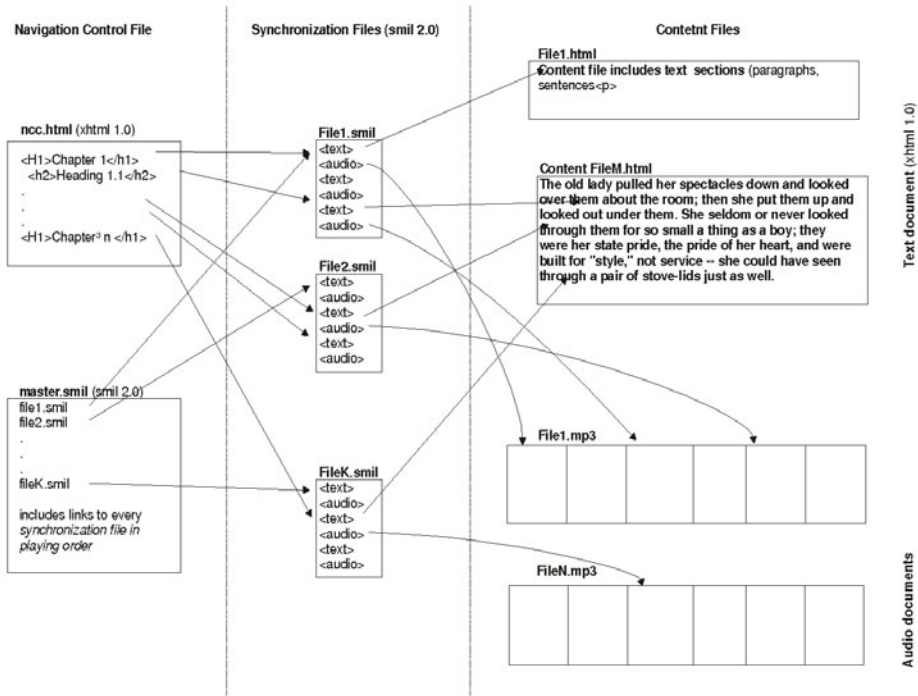


Fig. 1. Daisy book structure

3 Management System for Internet Multimedia Library

We present interactive system for online multimedia daisy books presentation, which main components are presented in Fig. 2 [4, 5]. The system is a result of an earlier research project conducted by The Silesian University of Technology together with the Academic Library and School for the Blind in Poland [2]. The design of the system allows:

- cataloguing and collecting of multimedia publications like e-books, e-magazines, digital talking books, digital music and movies
- assures secure Internet access to the library resources by registered users
- provides management of reader's orders
- distributing of the ordered publications on CDROM disks

The presented library system is running in Linux environment on Pentium multiprocessor servers. The system is managed with the web user interface and standard Internet web browsers. The user web interface was designed with the special attention to requirements of blind and low vision internet users and allows for direct access to information on internet pages, easy navigation and adjustment of font size, color and contrast as specified by the individual user.

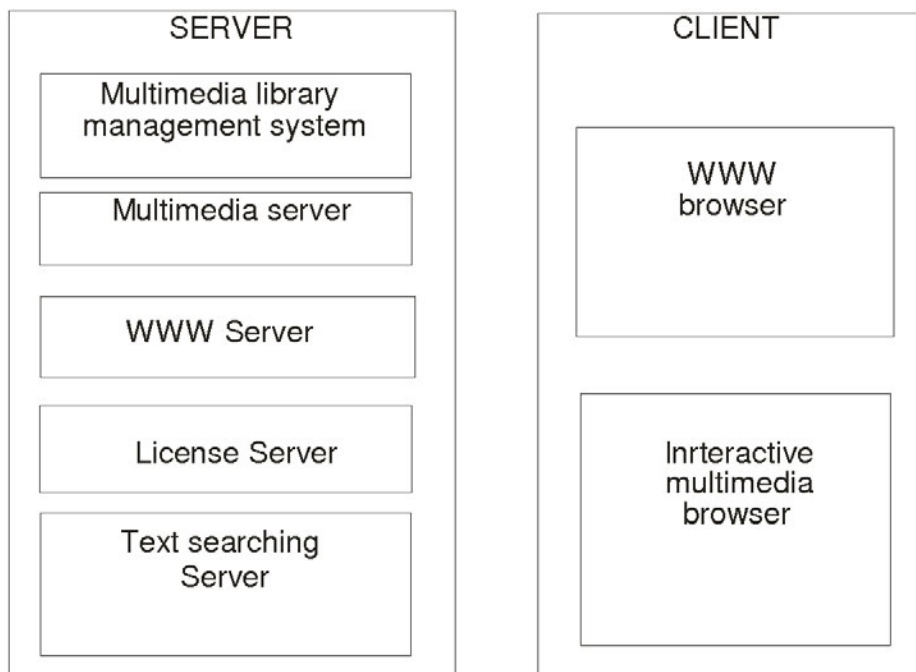


Fig. 2. Multimedia system components

WWW library service was build according to the W3C-WAI web content and section 508 accessibility guidelines.

4 Interactive Multimedia Daisy Book Browser

Currently available daisy players allow reading multimedia Daisy books to be stored on compact discs (CD) or local computer hard discs. Books recorded on CD-ROM's are collected personally or ordered by mail in the library. Some libraries offer digital books by Internet. This method requires full book contents download over the Internet. Books in Daisy format range from several to hundreds of megabytes in size. Downloading large amounts of data makes this approach to books' distribution both difficult and time consuming. Access to information contained in multiple books is very difficult and multiple books information searching is practically impossible. In our continuous research and development of the multimedia library system we have designed and developed new multimedia Daisy book browser. The new software Daisy reader allows playing Daisy books online over the Internet or in the standard way, from CD or from the local hard disk. Online Daisy books are played from a multimedia server and are available for reading after a few seconds from being found in the library system. Books audio, text and graphics content are presented synchronously.

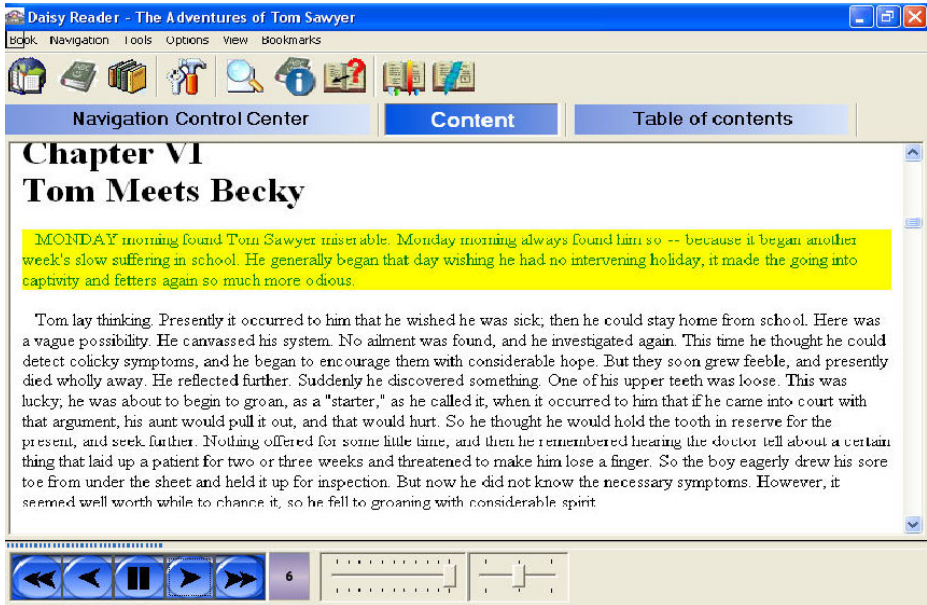


Fig. 3. Main window browser

Consecutive book pages are displayed. Sentences are highlighted with simultaneous audio being played. The book's index allows access to selected chapter. The reader can also navigate through the logical book structure by: chapters, headings, pages, tables, paragraphs and sentences. The user web interface was designed with the special attention to requirements of blind and low vision internet users and allows for direct access to information on internet pages, easy navigation and adjustment of font size, color and contrast as specified by the individual user. WWW library service was build according to the W3C-WAI web content and section 508 accessibility guidelines. The main window of the browser is presented in Fig. 3. Similar as in printed books readers can add bookmarks with text, audio notes and exchange bookmarks list each other. Browser offers searching text function which allows playing narrator's speech and presenting highlighted text from place where it has been founded. Browser user interface is customizable to different users group: blind, low vision, dyslectic, mobile impairment. Blind people can use browser with assistive software or use it in self voicing mode. We extend browser functionality which allows read aloud text DAISY book with synthetic voice. New navigation commands enable readers interactive audio browsing: sentence reading, word spelling and reading structural information like tables, math formulas [3]. Interactive audio presentation of math formulas and tables allows preparing advanced DAISY scientific, technical books and educational materials for students, mathematicians and scientists. Our research, which cooperates with DAISY MathML Project, lead to extends DAISY standard.

5 System Implementation

Our interactive multimedia Daisy book browser works together with multimedia Helix Universal server. Multimedia Helix server is integrated with the multimedia digital library management system. Using the web browser, users can search and browse library catalogue, after Daisy book selection system generates an encrypted license file (file with extension .dtb) this file contains access rights to the selected book. The web browser automatically starts the interactive multimedia Daisy book browser with the selected file. Next Daisy book browser establishes a connection with the streaming Helix Universal server. Text and graphics files and meta data describing book structure are accessed from WWW server. Multimedia book content in DAISY format is divided into parts and stored in many files: text xml, audio mp3 and graphics files. This is essential to continuous on-line book presentation over the Internet. Additional browser mechanisms preload book fragments and allow to present book contents without interruptions. Files are accessed in parallel with audio multimedia stream. Loaded xml files (containing book text) are buffered in Daisy browser memory which allows for smooth navigation. The user can search for information in the document before the document is loaded. The text search server implements this function. After the search request is processed and information about found text is passed back to browser. This information allows Daisy browser to playback book contents from any fragment that meets the search criteria. Selected multimedia streaming server (Helix Universal Server) works in a multiprocessor environment and allows for simultaneous data transmission and multi user service. During our work we tested the system's scalability with a dual processor Pentium IV server running Linux operating system. Scalability tests were conducted using special client and server applications with client applications count from 2 to 32. Test applications working together with an independent coordinating server allowed collecting various transmission quality statistics.

References

1. Spinczyk, D., Brzoza, P.: Online Internet Interactive System for Multimedia Daisy Books Presentation. *Accessible Design in the Digital World*, Dundee (2005)
2. Brzoza, P., Moroz, P.: Internet Multimedia Library Accessible to Impaired Readers. In: *ICEVI European Conference*, Chemnitz (2005)
3. Brzoza, P.: Presenting structural information in multimedia documents. In: *Computer Networks Conference*, Zakopane (2005)
4. Brzoza, P.: Presenting accessible information in multimedia environments. In: *Workshop New strategies for accessible information provision*, Paris (2004)
5. Brzoza, P.: Virtual multimedia library accessible to blind people. *Technology And Persons With Disabilities*, Los Angeles (2003)

Biomechanical Behaviour of Double Threaded Screw in Tibia Fixation

Witold Walke¹, Jan Marciniak¹, Zbigniew Paszenda¹, Marcin Kaczmarek¹,
and Jerzy Cieplak²

¹ Silesian University of Technology, Institute of Engineering Materials and Biomaterials, ul. Konarskiego 18a, 44-100 Gliwice, Poland
witold.walke@polsl.pl

² "BHH Mikromed", ul. Katowicka 11, 44-530 Dąbrowa Górnicza, Poland
info@mikromed.pl

Summary. The aim of the work was assessment of stability of tibia fixation realized with the use of double threaded screw. Biomechanical analysis of the tibia – double threaded screw system was carried out for the implant made of two biomaterials used in bone surgery – Cr-Ni-Mo stainless steel and Ti-6Al-4V alloy. Finite element method was applied to calculate displacements, strains and stresses. The obtained results allowed to work out biomechanical characteristics of the analyzed system. These characteristics can be a basis for selection of degree of strain hardening of the applied metallic biomaterial and optimization of the screw's geometry.

1 Introduction

Double threaded screws are new solution of tissue reconstructions in orthopaedics (stabilization of metaphysis fractures of long bones and spinal fractures) and dentistry. The most often the screws are applied in stabilizations of metacarpus and metatarsus fractures. The matter of these solutions is application of two threads of diverse diameter, assuring stabilization of bone fragments with the use of physiological effects [1, 2, 3, 4, 5, 6, 7].

Clinical experiences show that double threaded screws applied in orthopaedics and traumatology indicate many favorable features connected with both biomechanical quality of fixations and clinical results, especially with reference to minimization of tissue traumas.

The presented work is continuation of authors' research in the field of numerical analysis of diverse implants with the use of finite element method [8, 9, 10, 11, 12, 13, 14]. The main aim of the work was determination of biomechanical characteristics of a tibia – double threaded screw fixation.

2 Methods

Stabilization of an oblique tibia fracture with the use of the double threaded screw, mainly applied in phalangeal fixations, was analyzed in the work. On the

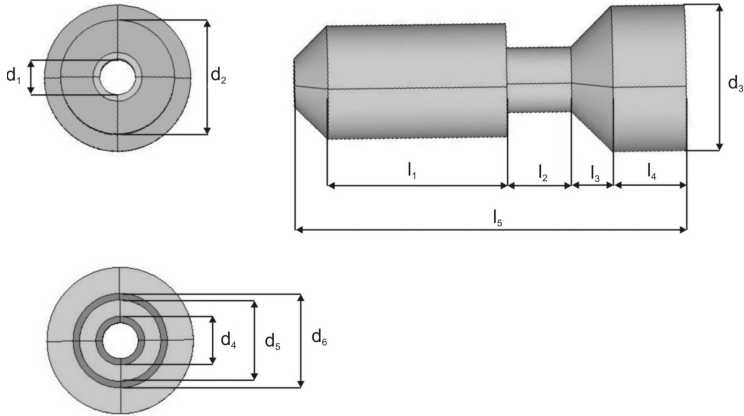


Fig. 1. Geometrical model of the double threaded screw

basis of anthropometric data, modification of screw's geometry was proposed – Fig. 1.

2.1 Numerical Model

Geometrical model of tibia was worked out on the basis of data collected from computer tomography of a real bone. The following parameters of tibia were established: Young's modulus $E=18600$ MPa and Poisson's ratio $\nu=0.4$ [15]. The geometrical model of the double threaded screw was worked out in ANSYS v.11. The following material properties were established:

- stainless steel – $E=2 \cdot 10^5$ MPa, Poisson's ratio $\nu=0.33$,
- Ti-6Al-4V alloy – $E=1.06 \cdot 10^5$ MPa, Poisson's ratio $\nu=0.33$.

Geometrical model of the tibia – double threaded screw system was presented in Fig. 2.

The geometrical models were discretized with the use of SOLID95 finite elements – Fig. 3. The analysis was carried out in order to calculate displacements, strains and stresses in:

- health tibia,
- elements of the system: tibia – double threaded screw made of stainless steel,
- elements of the system: tibia – double threaded screw made of Ti-6Al-4V alloy.

2.2 Boundary Conditions

In order to carry out calculations it was necessary to evaluate and establish initial and boundary conditions which imitate phenomena in real system with appropriate accuracy. The following assumptions were established:

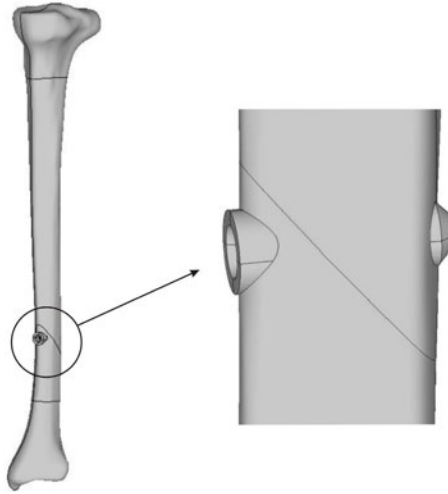


Fig. 2. Geometrical model of the tibia – double threaded screw system

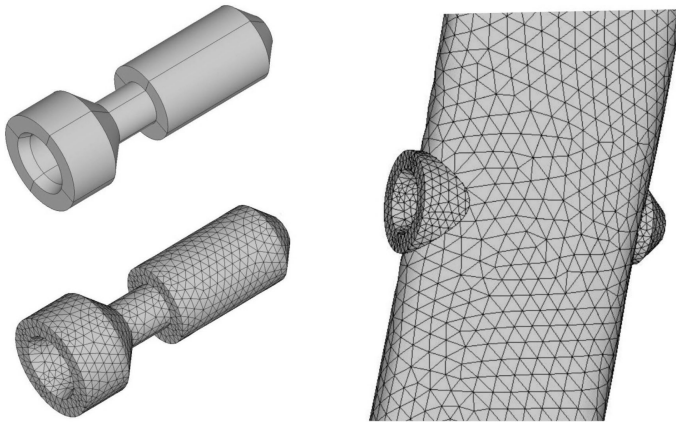


Fig. 3. Geometrical model meshed with SOLID95 elements

- distal fragment of tibia was immobilized (all degrees of freedom of surface nodes were taken away). It enabled displacements of the proximal fragment, blocking possible rotation,
- proximal fragment of the tibia was loaded with forces in the range $F=100\div 2000$ N with increment of 100 N,
- in the distal part of the tibia the oblique fractures was simulated (45°), enabling implantation of the double threaded screw according to the operating technique.

Stresses and strains obtained in the analysis are reduced values according to the Huber–Misses hypothesis.

3 Results

3.1 Tibia

The aim of this analysis was determination of influence of bone loading on stress distribution in the individual areas of the bone. Maximal stresses are localized in the distal, metaphysic part of the bone and for the maximal loading $F=2000\text{ N}$ are equal to $\sigma_{max}=66\text{ MPa}$. The obtained stresses did not exceed the strength of a bone ($\approx 160\text{ MPa}$) [15]. Example stress distribution in elements of the healthy bone, caused by the loading from the range $F=100\div 2000\text{ N}$ was presented in Fig. 4. The analysis of the healthy bone allowed to asses the area of maximum effort.

3.2 Tibia – Double Threaded Screw System

Results of the displacements, strains and stresses analysis carried out for the tibia – double threaded screw system were presented in Fig. 5–9. The analysis indicates that maximum displacements in the screw, calculated for diverse loading, were in the range $u=0.2\div 1.2\text{ mm}$ (for stainless steel) and $u=0.16\div 1.54\text{ mm}$ (for Ti-6Al-4V alloy) – Fig. 5.

Stress analysis showed that maximum reduced stresses were localized in the transition zone between threads (change of inner diameter) as well as in the area



Fig. 4. Stress distribution in health tibia for the applied loading: a) 100 N, b) 1000 N, c) 2000 N

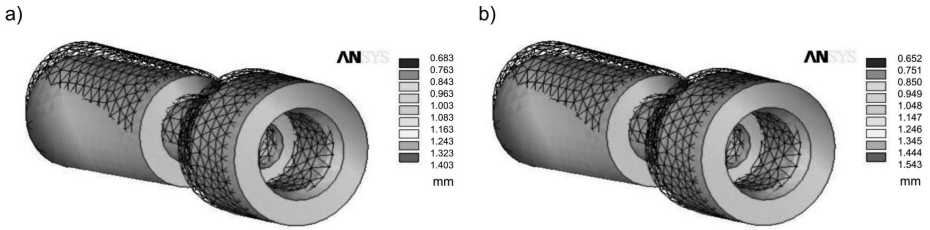


Fig. 5. Displacement distribution in the double threaded screw loaded with the force $F=2000\text{ N}$: a) stainless steel, b) Ti-6Al-4V alloy

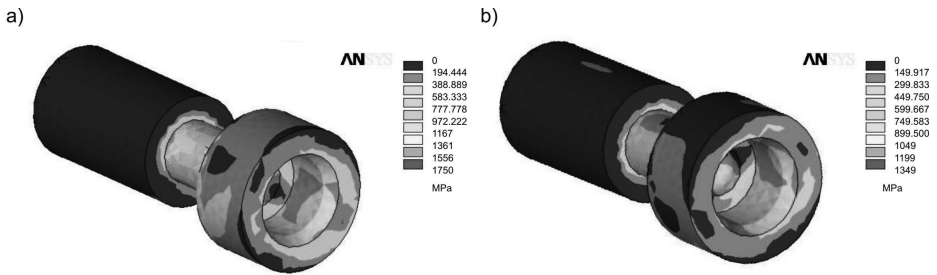


Fig. 6. Stress distribution in the double threaded screw loaded with the force $F = 2000\text{ N}$: a) stainless steel, b) Ti-6Al-4V alloy

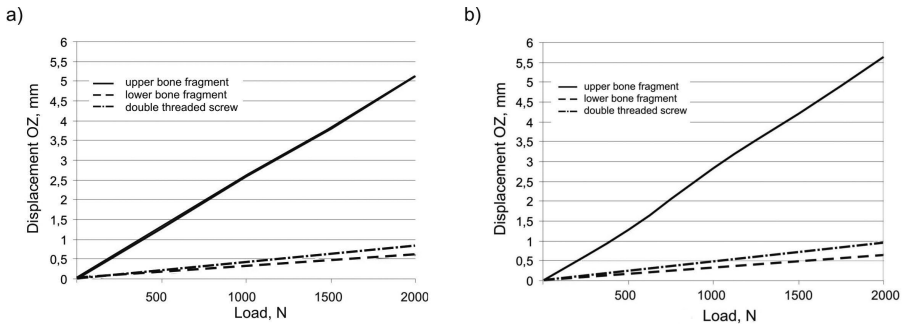


Fig. 7. Displacements in the OZ axis in a function of the applied loading: a) stainless steel, b) Ti-6Al-4V alloy

of direct contact between the bone and the screw. Values of reduced stresses, for the applied loading $F=100\div 2000\text{ N}$, were in the range $\sigma=2\div 1750\text{ MPa}$ (for stainless steel) and $\sigma=1\div 1321\text{ MPa}$ (for Ti-6Al-4V alloy) – Fig. 6. The maximum stresses were accompanied the maximum strains. The strains did not exceed the value of $\epsilon_{max}=0.89\%$ (for stainless steel) and $\epsilon_{max}=1.4\%$ (for Ti-6Al-4V alloy).

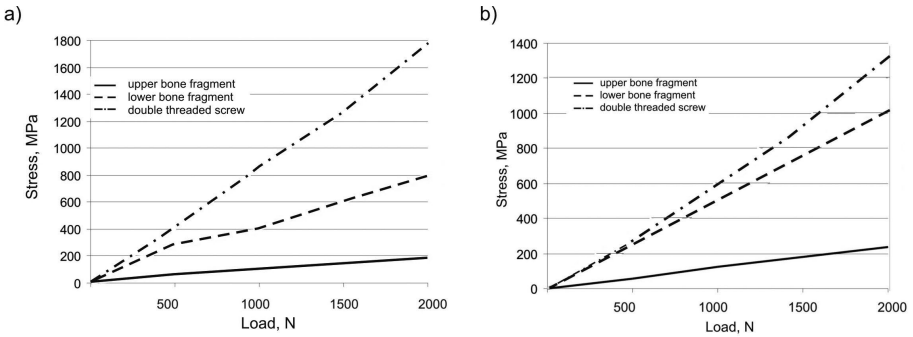


Fig. 8. Maximum stresses in a function of the applied loading: a) stainless steel, b) Ti-6Al-4V alloy

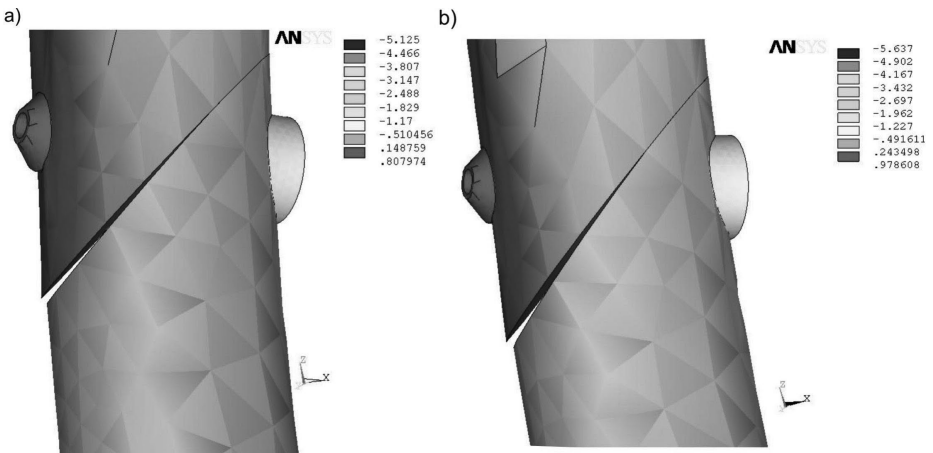


Fig. 9. Displacements in the fracture gap (OZ axis) for the maximum loading of 2000 N: a) stainless steel, b) Ti-6Al-4V alloy

The analysis allowed to work out biomechanical characteristics of the tibia – double threaded screw system. The characteristics present relation of displacements in the OZ axis and maximum reduced stresses in a function of the applied loading – Fig.7 and 8. Displacements in the fracture gap, determined along the OZ axis (bone axis) for the maximum loading $F=2000\text{ N}$, were presented in Fig. 9.

4 Conclusions

The aim of the work was assessment of stability of tibia fixation realized with the use of double threaded screw. Biomechanical analysis of the tibia – double threaded screw system was carried our for the implant made of two

biomaterials used in bone surgery – Cr-Ni-Mo austenitic stainless steel and Ti-6Al-4V alloy. Finite element method was applied to calculate displacements, strains and stresses. The oblique fracture of the tibia was localized in the distal part of the bone. This area was selected on the basis of the initial analysis of the health tibia – Fig. 4. The obtained results confirm clinical data about the most frequent fracture localizations.

In order to carry out calculations it was necessary to evaluate and establish initial and boundary conditions which imitate phenomena in real system with appropriate accuracy. The analyzed model was loaded with forces from the range $F=100\div 2000\text{ N}$. In fact, during stabilization of the fracture in time of rehabilitation such high loadings do not appear. The established range of forces purposed determination of the biomechanical characteristics in the widest possible range (from so called “biomechanical silence” – directly after the operation, to a physiological, dynamic loading).

The analysis of the tibia – double threaded screw system loaded with forces from the range $F=100\div 2000\text{ N}$ indicates that the obtained displacements are diverse – Fig. 5. It was affirmed that for the particular loading higher values of stresses are observed in the screw made of Ti-6Al-4V alloy. Together with the increase of the loading, greater difference in displacements for the analyzed biomaterials is observed.

The stresses in the tibia – double threaded screw system loaded with forces from the range $F=100\div 2000\text{ N}$ are diverse. The maximum values were equal to $\sigma=1750\text{ MPa}$ (for stainless steel) and $\sigma=1321\text{ MPa}$ (for Ti-6Al-4V alloy). Maximum reduced stresses were localized in the transition zone between threads – Fig. 6.

The obtained results allowed to determine biomechanical characteristics of the analyzed system ($u=f(F)$, $\sigma_{max}=f(F)$) – Fig. 7 and 8. The results are the basis for selection of degree of strain hardening of the applied metallic biomaterial and optimization of the screw’s geometry. Appropriate selection of mechanical properties and geometrical features of the implant is the main factor determining a stability of the fixation. Due to applied simplifications of the tibia – double threaded screw fixation model, the analysis results should be experimentally verified in laboratory conditions.

Acknowledgement. The work was supported by the research and development project no. R0801601 from Ministry of Science and Higher Education.

References

1. Brauer, R., Dierking, M., Werber, K.: Die Anwendung der Herbert-Schraube mit der Freehand-Methode zur Osteosynthese der frischen Skaphoidfraktur. Unfallchirurg 100, 776–781 (1997)
2. Herford, A., Ellis, E.: Use of a Locking Reconstruction Bone Plate/Screw System for Mandibular Surgery. Journal of Oral Maxillofacial Surgery 56, 1261–1265 (1998)

3. Lo, I., King, G., Milne, A., Johnson, J., Chess, D.: Biomechanical analysis of intrascaphoid compression using the Herbert scaphoid screw system. *Journal of Hand Surgery* 2, 209–213 (1998)
4. Faran, K., Ichioka, N., Trzeciak, M., Han, S.: Effect of bone quality on the forces generated by compression screws. *Journal of Biomechanics* 32, 861–864 (1999)
5. Lautenbach, M., Eisenschenk, A.: Aktueller stand der Kahnbeinchirurgie. *Trauma Berufskrankh* 4, 256–261 (2002)
6. DeVos, J., Vandenberghe, D.: Acute percutaneous scaphoid fixation using a non-cannulated Herbert screw. *Chirurgie de la main* 22, 78–83 (2003)
7. Sauerbier, M., Germann, G., Dacho, A.: Current Concepts in the Treatment of Scaphoid Fractures. *European Journal of Trauma* 2, 80–92 (2004)
8. Kajzer, W., Kaczmarek, M., Marciniak, J.: Biomechanical analysis of stent – oesophagus system. *Journal of Materials Processing Technology* 162-163, 196–202 (2005)
9. Walke, W., Paszenda, Z., Filipiak, J.: Experimental and numerical biomechanical analysis of vascular stent. *Journal of Materials Processing Technology* 164-165, 1263–1268 (2005)
10. Ziebowicz, A.: The use of miniplates in mandibular fractures – biomechanical analysis. *Journal of Materials Processing Technology* 175, 452–456 (2006)
11. Krauze, A., Marciniak, J.: Numerical method in biomechanical analysis of intramedullary osteosynthesis in children. *Journal of Achievements in Material and Manufacturing Engineering* 15, 120–126 (2006)
12. Walke, W., Paszenda, Z., Jurkiewicz, W.: Numerical analysis of three – layer vessel stent made form Cr-Ni-Mo steel and tantalum. *Journal of Computational Materials Science and Surface Engineering* 1, 129–139 (2007)
13. Kajzer, W.: Experimental and numerical analysis of urological stents. *Archives of Material Science and Engineering* 20, 297–300 (2007)
14. Krauze, A.: Numerical analysis of plates used in funnel chest treatment. *Engineering of Biomaterials* 67–68, 32–34 (2007)
15. Bedzinski, R.: *Engineering biomechanics*. Printing House of the Wroclaw University of Technology, Wroclaw (in Polish) (1997)

Biomechanical Analysis of Lumbar Spine Stabilization by Means of Transpedicular Stabilizer

Jan Marciniak¹, Janusz Szewczenko¹, Witold Walke¹, Marcin Basiaga¹, Marta Kiel¹, and Iлона Mańka²

¹ Silesian University of Technology, Institute of Engineering Materials and Biomaterials, ul. Konarskiego 18a, 44-100 Gliwice, Poland
marcin.basiaga@polsl.pl

² Silesian University of Technology, Department of Applied Mechanics, ul. Konarskiego 18a, 44-100 Gliwice, Poland

Summary. The fundamental purpose of research was determination of biomechanical characteristic of lumbar spine–transpedicular stabilizer system made of stainless steel (Cr-Ni-Mo) and Ti6Al4V alloy. To define biomechanical characteristic of the system finite element method was applied. Geometric models of part spine L3-L4 and stabilizer, was discretised by means of SOLID 95 element. Appropriate boundary conditions imitating phenomena in real system with appropriate accuracy were established. The aim of biomechanical analysis was calculation of displacements and stresses in the vertebrae and the stabilizer in a function of the applied loading: 700 N–1600 N. The results of the numerical analysis can be applied to determine a construction features of the stabilizer, and to select mechanical properties of metallic biomaterial. The defined displacements for vertebrae L3-L4 show that the proposed type of stabilizer enables correct course of treatment.

1 Introduction

Arthrosis of spine or overloadings cause damage of spine structures: vertebral segments, intervertebral discs or ligaments. Number of spine injuries with damage of spinal cord in Poland is estimated at the level of 600 to 800 annually. That includes road accidents (33–75%), falls from highs (12–44%), sport injuries (3.5–18%) and injuries of spinal cord [1].

Biomechanical problems of spine are not fully recognized. Knowledge of overloading causes and dysfunctions in consequence of instability determine further therapeutic management – both operative and rehabilitation. On the background of deformation causes, functional disorders and dysfunctions of spine, diverse stabilization systems can be applied.

Proposed solutions are not always followed by biomechanical analysis which allows to determine a characteristic connected with displacements in a function of applied loading. This fact is of great importance in risk assessment of the operation and the rehabilitation as well.

The most vulnerable part of spine, due to hyperostotic spondylitis, is a lumbar segment. There is located human's center of gravity and there also maximum forces loading vertebrae and intervertebral discs are observed. 62% of pathological changes in the vertebra–intervertebral disc system refers to the L3-L4 segment [2, 3].

Spinal instabilities are treated by means of diverse stabilization systems [4]-[13]. From the beginning of 80's the wide use of transpedicular screw systems is observed. Many producers of spine implants offer their individual solutions that differ in screw types and their assembly. The transpedicular stabilization system of spine enables treatment of thoracic, thoracic–lumbar and lumbar segment of spine. Geometric features of stabilizers' elements match individual anthropometric features of patients [3].

2 Materials and Methods

System of transpedicular stabilization of spine in the lumbar segment known from the patent [4] was analyzed in the work. The system consists of transpedicular screw, clamp element, nut, contact arm and supporting rod – Fig. 1.

Stabilization of two vertebrae of lumbar part was analyzed in the work – Fig. 2. Geometrical model of lumbar spine was prepared on the basis of data obtained from computer tomography of a real spine worked out in the Department of Applied Mechanics.

Biomechanical analysis of the of lumbar spine–transpedicular stabilizer system was realized with the model use of finite element method. The Ansys 11.0 program was applied. The implants properties were as follows:

- for Cr-Ni-Mo steel: $E=2 \cdot 10^5$ MPa, Poisson's ratio $\nu=0.33$,
- for Ti-6Al-4V alloy: $E=1.06 \cdot 10^5$ MPa, Poisson's ratio $\nu=0.33$.

Meshing was realized with the use of SOLID95 elements – Fig. 3.

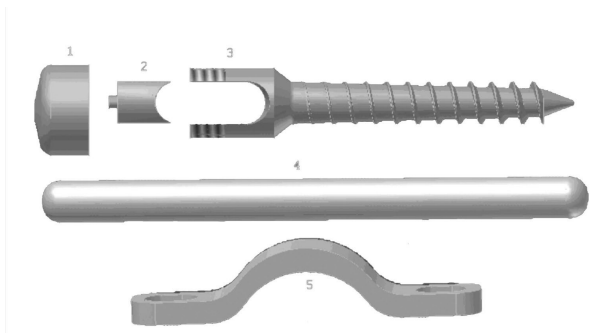


Fig. 1. Geometrical model of transpedicular stabilizer: 1 – nut, 2 – clamp element, 3 – screw, 4 – supporting rod, 5 – contact arm

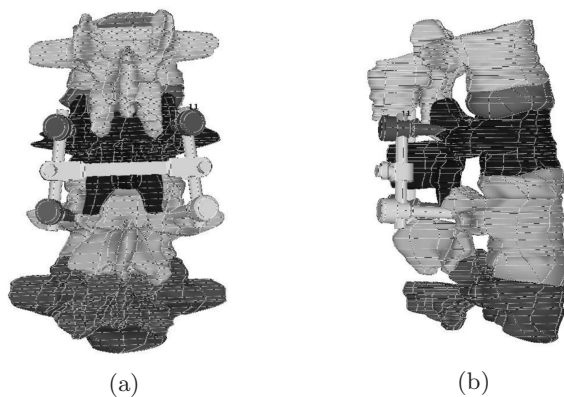


Fig. 2. Geometrical model of lumbar spine (L2-L5)–transpedicular stabilizer system

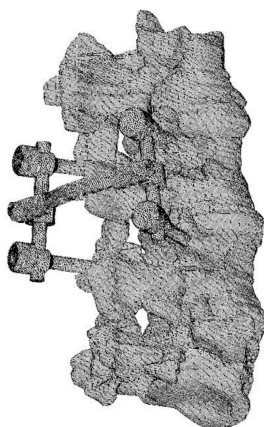


Fig. 3. Meshed model

In order to carry out calculations it was necessary to evaluate and establish initial and boundary conditions which imitate phenomena in real system with appropriate accuracy. The following assumptions were established:

- the fifth vertebra a part of lumbar spine was immobilized (all degrees of freedom of surface nodes were taken away). It enabled displacements at last lumbar vertebrae, blocking possible rotation,
- the second lumbar vertebra was loaded with forces: 700 N, 1000 N, 1300 N, 1600 N, on whole surface,
- in third and fourth vertebra the spine stabilizer was implanted according to the operating technique.

The scope of the analysis included determination of displacements and stresses:

- in the part of lumbar spine,
- in the vertebrae (L3-L4) – stabilizer system made of Cr-Ni-Mo,
- in the vertebrae (L3-L4) – stabilizer system made of Ti-6Al-4V alloy.

Stresses and strains obtained in the analysis are reduced values according to the Huber – Misses hypothesis.

3 Results

Results of the analysis carried out for the part of lumbar spine–transpedicular stabilizer system (made of Cr-Ni-Mo steel) are presented in Table 1, 2 and Fig. 4, 5.

On the basis of the analysis it was affirmed that maximum displacements was 0.34mm for the forces of 1600 N. However it was affirmed that maximum stress

Table 1. Results of the analysis displacements of spine–transpedicular stabilizer system made of Cr-Ni-Mo steel

Displacements, mm					
F,N	Maximum	OZ	OY	OX	
700	0.14	0.02	0.01	0.025	
1000	0.21	0.045	0.013	0.048	
1300	0.28	0.053	0.02	0.051	
1600	0.34	0.062	0.020	0.056	

Table 2. Results of the analysis stresses of spine–transpedicular stabilizer system made of Cr-Ni-Mo steel

Stresses σ , MPa						
F,N	Contact arm	Screw	Vertebrae	Intervertebral disc	Maximum	
700	28.2	33	9	2	32	
1000	40.32	48.1	12.11	3	46	
1300	52.1	62.5	16.2	3.1	60	
1600	63.56	76.42	19.21	3.58	74	

Table 3. Results of the analysis displacements of spine–transpedicular stabilizer system made of Ti6Al4V alloy

Displacements, mm					
F,N	Maximum	OZ	OY	OX	
700	0.15	0.02	0.007	0.02	
1000	0.21	0.032	0.01	0.045	
1300	0.28	0.04	0.012	0.051	
1600	0.34	0.064	0.016	0.056	

for the same force was in the left top transpedicular screw. The stress reached 77 MPa – Fig. 4, 5 and Table 1, 2.

Results of the analysis carried out for the part of lumbar spine–transpedicular stabilizer system (made of Ti6Al4V alloy) are presented in Table 3, 4 and Fig. 6, 7.

On the basis of the analysis it was affirmed that maximum displacements was 0.34 mm for the force of 1600 N. However it was affirmed that maximum stress – 55 MPa – for the same force was localized in the left top transpedicular screw – Fig. 6, 7 and Table 3, 4.

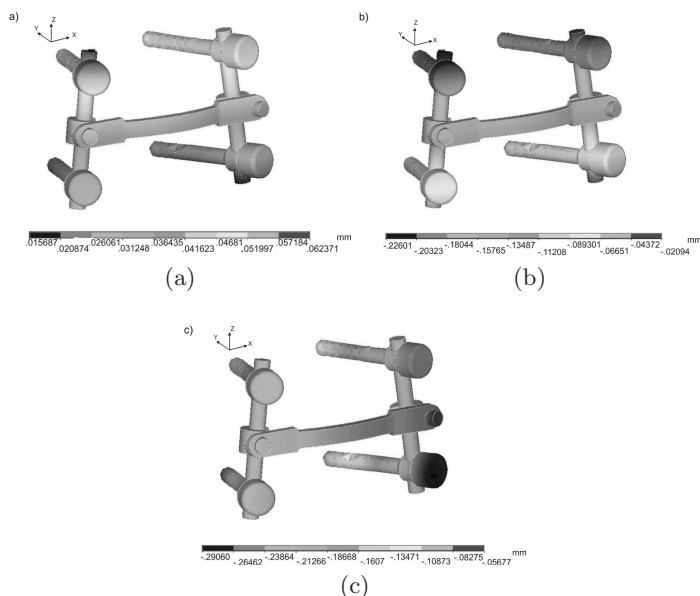


Fig. 4. Displacements in the spine of stabilizer loaded with the force 1600N (Cr-Ni-Mo steel) a) OZ axis, b) OY axis, c) OX axis

Table 4. Results of the analysis stresses of spine–transpedicular stabilizer system made of Ti6Al4V alloy

F,N	Stresses σ , MPa				
	Contact	arm	Screw	Vertebras	Intervertebral disc
700	19.42	23.75	11.3	1.68	32.44
1000	27.8	34	16.18	2.41	46.44
1300	36.37	44.48	21.16	3.15	60.75
1600	44.56	54.49	25.93	3.86	74.43

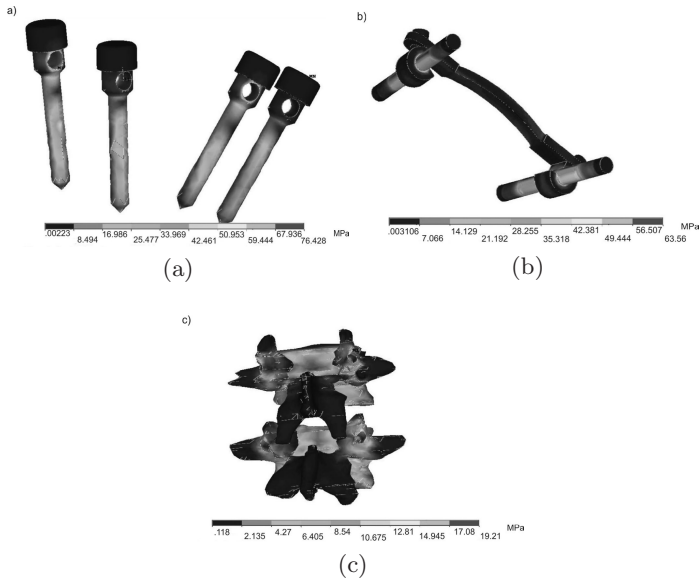


Fig. 5. Stresses loaded with the force 1600 N (Cr-Ni-Mo steel) a) transpedicular screw, b) contact arm, c) vertebrae (L3-L4)

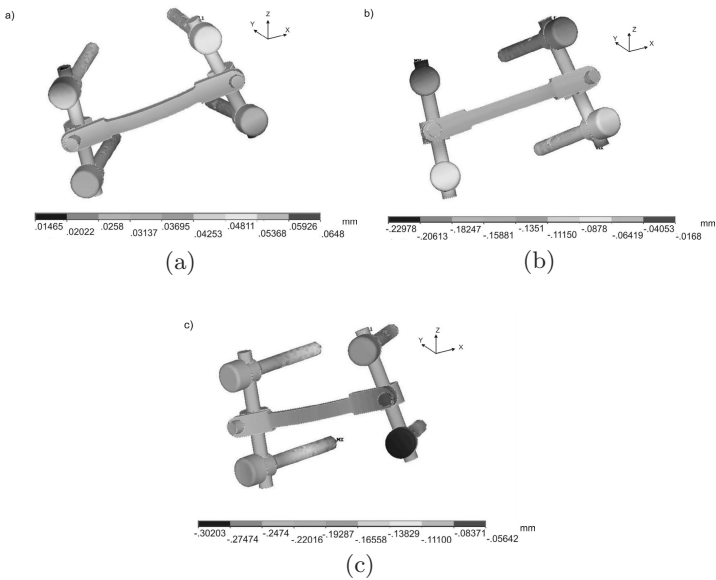


Fig. 6. Displacements in the spine of stabilizer loaded with the force 1600 N (Ti6Al4V alloy) a) OZ axis, b) OY axis, c) OX axis

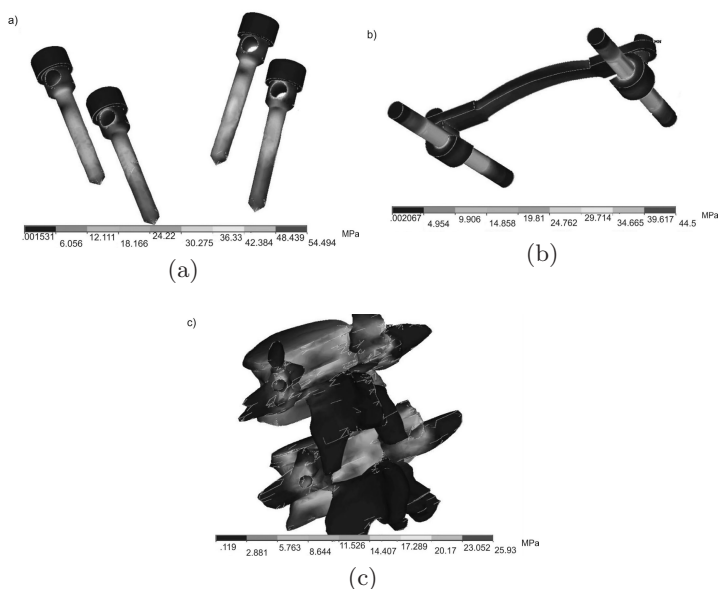


Fig. 7. Stresses loaded with the force 1600 N (Ti6Al4V alloy) a) transpedicular screw, b) contact arm, c) vertebrae (L3-L4)

4 Conclusions

The work presents results of biomechanical analysis of the part spine–transpedicular stabilizer system. The analysis was carried out with the use of finite element method. Displacements and stresses in the system’s elements were calculated. In work presented the most often using metallic biomaterials – Cr-Ni-Mo steel and Ti6Al4V alloy. Susceptibility of the system to displacements caused by the applied loading is the important parameter influencing the effectiveness of the proposed stabilization. Therefore the displacements between L3-L4 vertebrae, for the maximum loading of 1600N, were calculated – Fig. 4, 6. Depending on the applied material, no significant differences in displacements between vertebrae in the OZ were observed – Table 1 and 3. For the stabilizer made of with Cr-Ni-Mo steel and Ti6Al4V alloy displacements did not exceed 0.1 mm. Assembly of the stabilizer in L3-L4 segment and applying the maximum force of 1600 N does not cause the stress increase in the vertebral segments exceeding the value of 160 MPa.

Acknowledgement. The work was supported by the research and development project no. R0801601 from Ministry of Science and Higher Education.

References

1. Kiwerski, J.: Schorzenia i urazy kręgosłupa. PZWL, Warszawa (in Polish) (2004)
2. Nabrani, F., Wake, M.: Komputer modelling and stress analysis of the lumbar spine. *Materials Processing Technology* 127, 40–47 (2002)
3. Będziński, R.: *Biomechanika inżynierska. Zagadnienia wybrane*. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław, 13–45 (in Polish) (1997)
4. Będziński, R., Filipiak, J., Pezowicz, C., Marciniak, J.: Stabilizacja transpedikularna kręgosłupa do leczenia złamań i zniekształceń, Patent nr. 3356/29/03 (in Polish) (2003)
5. Doherty, B., Heggenes, M.: The quantitative anatomy of the atlas. *Spine* 19(22), 2497–2501 (1994); COMMENT 2005 of *Materials Processing Technology* 162–163, 209–214 (2005)
6. Marciniak, J.: Austenitic steel – the basic implantation material used in orthopaedic surgery. *Ortopedia, Traumatologia, Rehabilitacja* 3, 52–58 (in Polish) (2000)
7. Marciniak, J.: *Biomaterials*, Edit by Silesian Univesity of Technology pp. 116, 219–229, 238–252, Gliwice (in Polish) (2002)
8. Marciniak, J.: Perspectives of employing of the metallic biomaterials in the reconstruction surgery. *Engineering of Biomaterials* (1), 12–20 (1997)
9. Fantigrossi, A., Galbusera, F., Raimondi, M.: Biomechanical analysis of cages for posteriori lumbar interbody fusion. *Medical Engineering & Physics* 29, 101–109 (2007)
10. ISO 4967 Norm (E). Steel – Micrographic determination of content of non–metallic inclusions –Micrographic method using (1979)
11. Mc Lain, M., Fry, M.: Lumbar pedicle screw salvage: putton testing of thres different pedicle screw designs. *Spinal Disorders* 8, 62–65 (1995)
12. Paszenda, Z., Tyrlik-Held, J., Marciniak, J., Włodarczyk, A.: Corrosion resistance of Cr-Ni-Mo steel intended for implants used in operative cardiology. In: *Proceedings of the 9th International Scientific Conference Achievements in Mechanical and Materials Engineering 2000*, Gliwice-Sopot-Gdańsk, pp. 425–428 (2000)
13. PN - ISO 5832–1, Implants for surgery metallic materials, Part I: Wrought stainless steel (1997)

FEM Analysis of the Expandable Intramedullary Nail

Wojciech Kajzer, Anita Krauze, Marcin Kaczmarek, and Jan Marciniak

Institute of Engineering Materials and Biomaterials, Silesian University of Technology
{wojciech.kajzer, anita.krauze, marcin.kaczmarek, jan.marciniak}@polsl.pl

Summary. The paper presents results of numerical analysis of new form of expandable intramedullary nail (patent no. P382247) used in stabilization of proximal femur in adults. The obtained results can be used to optimize geometry of implants as well as mechanical properties of metallic biomaterial they are to be made of.

1 Introduction

Biomechanical quality of a bone - intramedullary nail fixation is important issue of remodeling in nailing osteosynthesis. The biomechanical analysis can be carried out for selected model, construction of the nail and its fastening. On the basis of biomechanical analyses, both geometry and mechanical properties of biomaterial as well as physico-chemical properties can be formed. Biomechanical characteristics of nails enable to compare and select a stabilization method for individual patients.

Nowadays, elastic methods of osteosynthesis are promoted. The basic aim of these methods is assuring micromovements of bone fragments that stimulate remodeling of bone by differentiation of its structure. Strains in bone tissue in the elastic range generate electromechanical potentials in bone. Therefore, establishing the optimal axial, transverse and torsional stiffness is crucial.

Determination of stresses and strains in intramedullary osteosynthesis can be applied in selection of mechanical properties of nails biomaterial and in forming of structure and physio-chemical properties of surface as well. Biocompatibility of implants is considered with reference to metabolic, bacteriological, immunological and oncogenic processes. It is connected with individual reactivity of implants' user. Therefore, biomaterials of even identical mechanical properties but diverse physio-chemical properties of surface should be differentiated [1]–[8].

2 Materials and Methods

Numerical model of femur, worked out in Laboratorio di Tecnologia dei Materiali, Istituti Ortopedici Rizzoli, was applied in the biomechanical analysis of the expandable intramedullary nail. Young's modulus $E=18600$ MPa and Poisson's ratio $\nu=0.3$ were assumed for femur model.

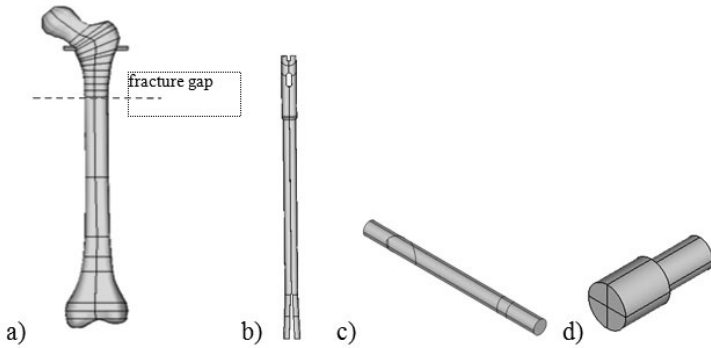


Fig. 1. Geometrical model of the femur – expandable intramedullary nail system: a) view of the system, b) expandable intramedullary nail, c) lock, d) blocking screw

Geometrical model of expandable intramedullary nail was prepared in ANSYS. The following mechanical properties were selected:

- Stainless steel Cr-Ni-Mo: $E=2 \cdot 10^5$ MPa, Poisson's ratio $\nu=0.33$
- Ti-6Al-4V alloy: $E=1.1 \cdot 10^5$ MPa, Poisson's ratio $\nu=0.33$.

Geometrical model of the analyzed femur - expandable intramedullary nail system was presented in Fig. 1. The analysis was carried out for proximal simple fracture (100 mm below trochanter) – Fig. 1. On the basis of the geometrical models a finite element mesh was generated – fig. 2a. The meshing was realized with the use of the SOLID95 element – Fig. 2b. This type of element is used for the three-dimensional modeling of solid structures. The element is defined

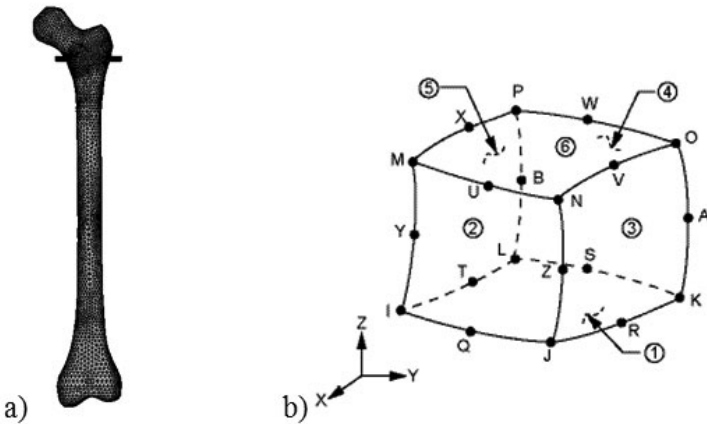


Fig. 2. a) Discrete model of the femur – expandable intramedullary nail system, b) The SOLID 95 finite element

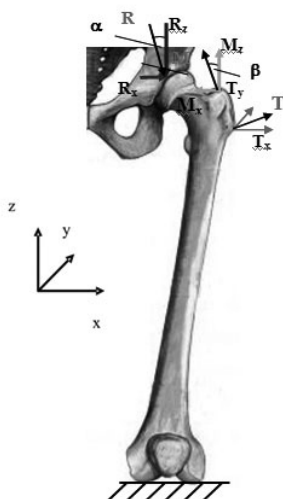


Fig. 3. Loading scheme of model

by eight nodes having three degrees of freedom at each node: translations in the nodal x , y , and z directions.

In the course of the work, displacements, strains and stresses, depending on the assumed mechanical properties, were calculated. In order to carry out the calculations, appropriate initial and boundary conditions reflecting phenomena in real system were determined. The following assumptions were set:

- lower part of the femur was immobilized (all degrees of freedom of nodes on external surfaces of condyles were taken away),
- proximal part of femur was loaded according to the scheme presented in fig. 3. The applied loading was presented in table 1.

The first stage of the analysis was determination of displacements, strains and stresses:

- in healthy femur,
- in elements of the femur – expandable intramedullary nail made of stainless steel,
- in elements of the femur – expandable intramedullary nail made of Ti-6Al-4V alloy.

Table 1. Forces applied to the femur [1]

Forces, N								
R			M			T		
x	y	z	x	y	z	x	y	z
494	-1824	0	-494	1208	0	-54	-21	0

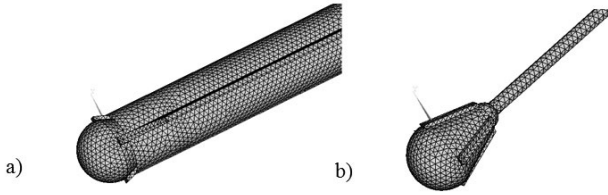


Fig. 4. Numerical model of the expandable part of the intramedullary nail – a) and expander – b) after discretization with SOLID 95 finite element

The obtained stresses and strains were reduced values according to the Huber-Misses-Henck hypothesis.

The second stage of the work was analysis of the nail during expansion. Axial displacement of the expander from 3 mm (contact with expandable end of the nail) to 7 mm with increment equal to 1 mm was analyzed. In order to carry out the analysis geometrical model of the expander and the expandable end of the nail were discretized by SOLID 95 element – Fig. 4.

Calculations were carried out in ANSYS 11 with the use of PC of the following parameters: Procesor Intel Core 2 Duo E6600, 4 GB RAM, Windows Vista Ultimate 64 bit.

3 Results

3.1 Results of the Femur – Expandable Intramedullary Nail System Analysis

The maximum obtained values of displacements, strains and stresses for all analyzed variants were presented in table 2 and Fig. 5, 6, 7.

Table 2. Results of the FEM analysis of the femur – intramedullary nail system

	Displacement. mm				Strains ϵ . %				Strains σ . MPa			
	x	y	z	Σ	x	y	z	Σ	x	y	z	Σ
Femur												
Femur	-15.8	0.5	1.8	16.2	10	3	19	54	222	148	452	635
Femur – expandable intramedullar nail system (Cr-Ni-Mo steel)												
Femur	-16.5	0.4	3.2	17.0	19	7	13	38	589	281	404	706
Nail	-14.1	0.3	3.2	14.1	2	2	9	18	2014	2030	4594	2899
System	-16.5	0.4	3.2	17.0	19	7	21	38	2680	2920	5713	4332
Femur – eexpandable intramedullar nail system (Ti-6Al-4V alloy)												
Femur	-20.4	0.8	4.2	21.1	18	8	13	46	475	303	471	866
Nail	-17.3	0.6	4.2	17.3	4	2	16	29	1765	1210	2708	2844
System	-20.4	0.8	4.2	21.1	18	8	37	46	2714	2818	5670	3938

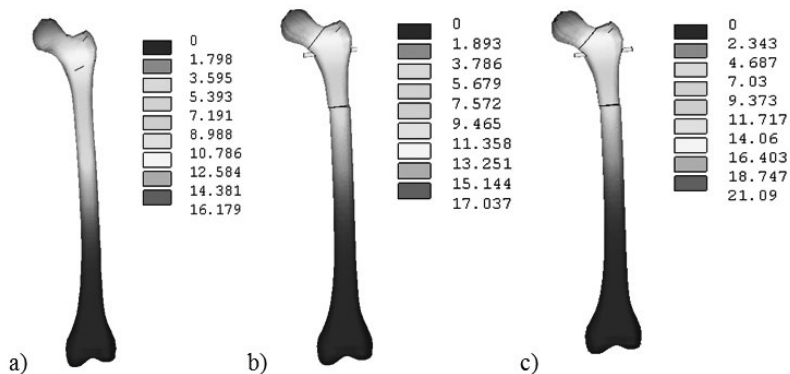


Fig. 5. Displacement vector sum, mm a) femur, b) femur – intramedullary nail system (Cr-Ni-Mo), c) femur – intramedullary nail system (Ti-6Al-4V)

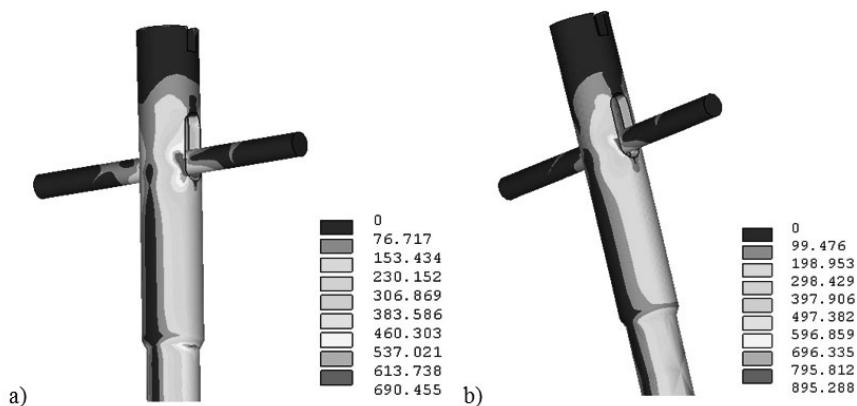


Fig. 6. Stress distribution in the nail, MPa: a) (Cr-Ni-Mo) nail, c) (Ti-6Al-4V) nail

The analysis showed no significant differences in displacements of the head for the healthy bone and the bone with the implanted nail. For the system with the nail made of stainless steel, the displacement of femoral head was equal to 17.0 mm. However, for the system with the nail made of titanium alloy, the displacement was equal to 21.1 mm. This indicates stiffness comparability of the healthy bone to the bone with the implanted nail – Fig. 5.

Maximum reduced stresses were localized in the area of contact between the lock and the nail. In the contact point the maximum value was equal to 4332 MPa for the stainless steel and 3938 MPa for the titanium alloy. But the analysis of the whole nail indicates that stresses did not exceed 690 MPa for the steel and 895 MPa for the titanium alloy – Fig. 6.

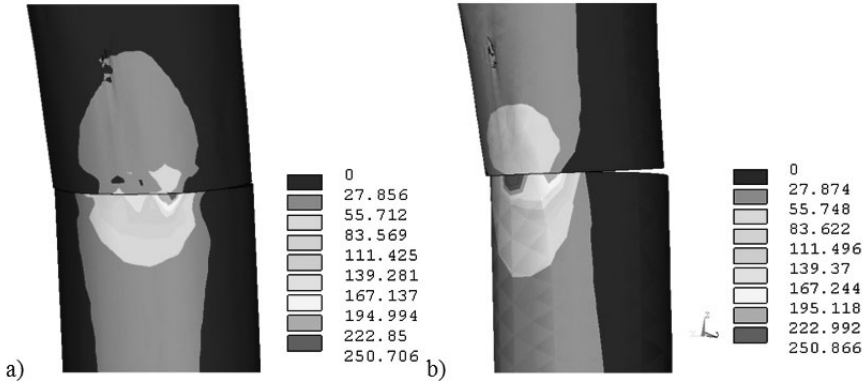


Fig. 7. Stress distribution in the fracture gap, MPa: a) (Cr-Ni-Mo) nail, c) (Ti-6Al-4V) nail

Also stresses in the fracture gap only locally exceeded the allowable value (250 MPa). Exceeding of the value causes damage of bone tissue. Maximum stresses are localized in the area of contact between the bone and the lock. On the basis of clinical research it was affirmed that bone is characterized by visco-elastic properties which allow to adapt tissues to existing loading without damage.

3.2 Results of the Expansion of the Intramedullary Nail Analysis

In the result of calculations, displacements, strains and reduced stresses were determined. Furthermore, characteristics of the expander and the expanding part of the intramedullary nail were also worked out – Table 3 and Fig. 8 and Fig. 9.

Table 3. Results of the FEM analysis of the expandable intramedullary nail expansion

	Axial displacement of the expander				
	3	4	5	6	7
Cr-Ni-Mo steel					
Displacement of expandable end of nail r. mm	0.52	1.03	1.50	2.02	2.48
Strains ϵ . %	0.18	0.36	0.53	0.71	0.82
Stresses σ . MPa	365	722	1039	1408	1721
Ti-6Al-4V alloy					
Displacement of expandable end of nail r. mm	0.52	1.03	1.50	2.02	2.48
Strains ϵ . %	0.18	0.36	0.53	0.71	0.82
Stresses σ . MPa	199	397	571	774	946

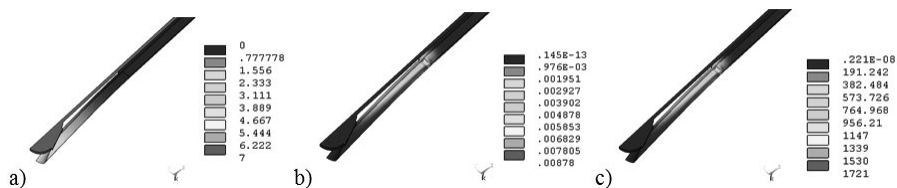


Fig. 8. Results of the analysis for the expander’s displacement equal to 7 mm for the nail made of stainless steel: a) displacements of the expander and the expandable end, mm, b) reduced strains in the expandable end, x100%, c) reduced stresses in the expandable end, MPa

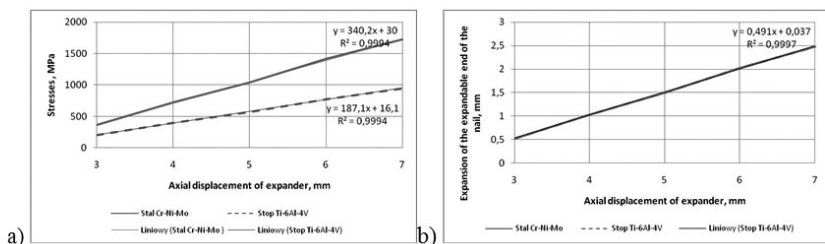


Fig. 9. a) stresses in the expandable end in a function of axial displacements of the expander, b) expansion of the expanding end in a function of axial displacements of the expander

The analysis showed that axial displacement of the expander is accompanied by linear increase of strains and reduced stresses up to maximum values. For the axial displacement equal to 7 mm maximum stresses for the steel are equal to $\sigma = 1721$ MPa and for the Ti-6Al-4V alloy $\sigma = 946$ MPa – Fig. 8a, b, c and Fig. 9a.

Independently on the applied biomaterial, the expansion of the end was the same and increased linearly depending on the axial displacement of the expander, reaching the maximum value $r = 2.48$ mm – Fig. 9b.

4 Conclusion

The numerical analysis was carried out in order to calculate displacements, strains and stresses in the expandable intramedullary nail used in stabilization of proximal femur in adults. The obtained results are important for selection of mechanical properties of metallic biomaterials intended for this type of implants. On the basis of the established boundary conditions and obtained results one can conclude that:

- maximum reduced stresses were localized in the area of contact between the lock and the nail. In the contact point the maximum value was equal to

4332 MPa for the stainless steel and 3938 MPa for the titanium alloy. But the analysis of the whole nail indicates that stresses did not exceed 690 MPa for the steel and 895 MPa for the titanium alloy,

- also stresses in the fracture gap, contact area of the bone fragments as well as in the whole bone did not exceed the allowable value of 250 MPa. Exceeding of the value causes damage of bone tissue,
- the analysis showed no significant differences in displacements of the head for the healthy bone and the bone with the implanted nail. For the system with the nail made of stainless steel, the displacement of femoral head was equal to 17.0 mm. However, for the system with the nail made of titanium alloy, the displacement was equal to 21.1 mm. This indicates stiffness comparability of the healthy bone to the bone with the implanted nail,
- for the proposed geometry of the nail the allowable expansion of the expandable end was equal to 1.03 mm for the stainless steel (axial displacement of the expander was equal to 4 mm) and 2.48 mm for the titanium alloy (axial displacement of the expander was equal to 7 mm).

Acknowledgement. The work was realized within the confines of the research project MNiSzW Nr R08 016 01 funded by the Minister of Science and Information Society Technologies.

References

1. Marciniak, J., Chrzanowski, W., Krauze, A., Kajan, E.: Gwoździowanie śródszpikowe w osteosyn-tezie. Wyd. Politechniki Śląskiej, Gliwice (in Polish) (2006)
2. Marciniak, J.: Biomateriały. Wyd. Politechniki Śląskiej, Gliwice (in Polish) (2002)
3. Paszenda, Z.: Problematyka tworzyw metalowych stosowanych na implanty w kardiologii zabiegowej. *Inżynieria Biomateriałów* 21, 3–9 (in Polish) (2002)
4. Chrzanowski, W., Marciniak, J.: Biomechanical analysis of the femoral bone-interlocking intramedullary nail system. In: 18th European Conference on Biomaterials, Stuttgart, Germany (2003)
5. Marciniak, J., Chrzanowski, W., Kaczmarek, M. Biomechaniczna analiza układu kość udowa-gwóźdź śródszpikowy z wykorzystaniem metody elementów skończonych. *Inżynieria Biomateriałów*, 30–33, 53–55 (in Polish) (2003)
6. Chrzanowski, W., Marciniak, J.: Displacement analysis in the femoral bone – intramedullary locked nail system. In: Proceedings of The Congress European Society of Biomechanics ESB 2004, S-Hertogenbosch, Netherlands (2004)
7. Krauze, A., Marciniak, J.: Biomechanical analysis of a femur-intramedullary nails system in children. In: 22nd DAS - 2005, Danubia-Adria Symposium on Experimental Methods and Solid Mechanics, Parma, Italy, 28.09–01.10 (2005)
8. Krauze, A., Marciniak, J.: Numerical method in biomechanical analysis of intramedullary osteosynthesis in children. In: Proc. of 11th Int. Sci. Conf. CAM3S 2005 Contemporary Achievements in Mechanics, Manufacturing and Materials Science, Gliwice-Zakopane, pp. 528–533 December 6–9 (2005)

Biomechanical Analysis of Plate for Corrective Osteotomy of Tibia

Jan Marciniak¹, Marcin Kaczmarek¹, Witold Walke, and Jerzy Cieplak²

¹ Silesian University of Technology, Institute of Engineering Materials and Biomaterials, ul. Konarskiego 18a, 44-100 Gliwice, Poland

witold.walke@polsl.pl

² "BHH Mikromed", ul. Katowicka 11, 44-530 Dąbrowa Górnicza, Poland

info@mikromed.pl

Summary. The aim of the work was assessment of system for corrective osteotomy of tibia (patent no. P382316). The system consisted of the plate of shape adapted to anatomical curvature of bone and the distance block, assembled together with the plate by means of connective screws. Biomechanical analysis of the tibia – plate system was carried out for the implant made of two biomaterials used in bone surgery – stainless steel and Ti-6Al-4V alloy. Finite element method was applied to calculate displacements, strains and stresses. The obtained results allowed to work out biomechanical characteristics of the analyzed system. These characteristics can be a basis for selection of degree of strain hardening of the applied metallic biomaterial and optimization of the plate's geometry.

1 Introduction

Indications for tibia osteotomies are valgus or varus deformities of knees. Qualification for osteotomy are: flexions at least to the angle of 90°, contracture in flexion less than 15-20°, joint stability. Clinical research on the most effective corrections of knee joint axis allowed to work out many types of osteotomies of proximal tibia epiphysis. They can be divided into several groups: linear, cuneiform, geometrical and hinge osteotomies [1].

In the 60's a plate osteosynthesis according to the AO was introduced. Since the very beginning this method had many disadvantages, and the most important one was an osteolysis in a contact area. The another disadvantage was a bone atrophy in the fracture site. A large number of complications was reported [1, 2]. That's why the AO method should be numbered among the over-rigid methods of stabilization, which lead to the demineralization of the bone tissue and the loss of the mechanical properties in consequence. The over-rigid stabilization also leads to the stress increase in the elements of the stabilizer which can cause the damage of the stabilizer. Furthermore, the initiation of cracks in the plate can occur, increasing the danger of the metalosis and bone union complications. The osteolysis and the osteoporosis observed in plate fixations are important problems [4]-[9]. The research concern mechanical properties of animal bones in the different stages of the fracture healing [10, 11] and the bone structure [12].

In the 70's the research on the elastic fixations with the use of the silicon pads [13], plastic plates [14] and autocompressing screws [15] have appeared.

Biophysical processes in bones and their properties as a mechanoreceptor are the justification for elastic methods of osteosynthesis. The methods guarantee the possibility of cyclic elastic strains of the bone while loading, so they activate the bone union. These methods of stabilization are recently appreciated in a clinical practice. Elastic stabilizers are the subject of research in many centers all over the world.

2 Materials and Methods

Geometrical model of tibia was worked out on the basis of data collected from computer tomography of a real bone. The following parameters of tibia were established: Young's modulus $E=18600$ MPa and Poisson's ratio $\nu=0.4$. The geometrical model of the plate was worked out in ANSYS v.11. The following material properties were established:

- stainless steel - $E=2 \cdot 10^5$ MPa, Poisson's ratio $\nu=0.33$
- Ti-6Al-4V alloy - $E=1.06 \cdot 10^5$ MPa, Poisson's ratio $\nu=0.33$

Geometrical model of the tibia – plate system was presented in Fig. 1a and b.

The geometrical models were meshed with the use of SOLID95 finite elements – Fig. 1c. This type of element is characterized by 20 nodes and 3 degrees of freedom in each node (displacements in x, y and z direction).

The analysis was carried out in order to calculate displacements, strains and stresses, depending on the applied mechanical properties of the plate. In order to carry out the calculations it was necessary to evaluate and establish initial and

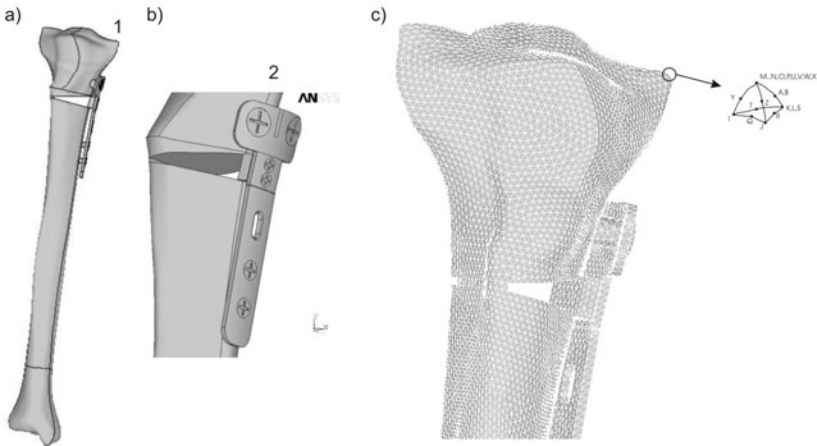


Fig. 1. Geometrical model of the tibia – plate system: a) general view, b) implanted system, c) meshed model (SOLID95)

boundary conditions which imitate phenomena in real system with appropriate accuracy. The following assumptions were established:

- distal fragment of tibia was immobilized (all degrees of freedom of surface nodes were taken away). It enabled displacements of the proximal fragment, blocking possible rotation,
- upper epiphysis of the tibia was loaded with forces: 100 N, 500 N, 1000 N, 1500 N, 2000 N,
- proximal segment of upper epiphysis was osteotomized according to the operating technique.

The analysis was carried out in order to calculate displacements, strains and stresses in:

- elements of the tibia – stainless steel plate system,
- elements of the tibia – Ti-6Al-4V plate system.

Stresses and strains obtained in the analysis are reduced values according to the Huber – Misses hypothesis.

3 Results

3.1 Results of Tibia – Stainless Steel Plate System

Analysis showed that maximum stresses, for the applied maximum loading $F=2000$ N, were localized in the screw implanted to the proximal tibia's shaft – Fig. 2b.

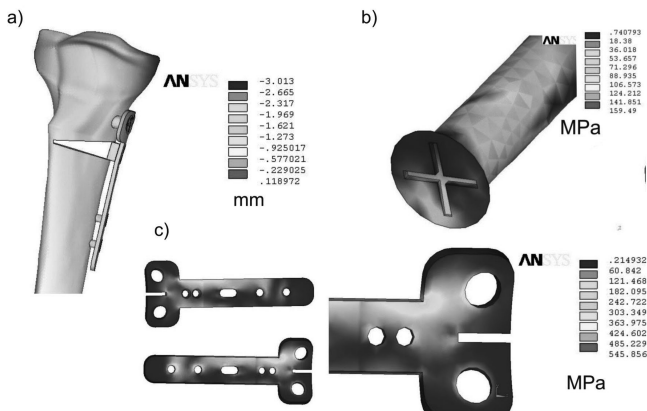


Fig. 2. Results of biomechanical analysis of the bone – plate system (the applied loading: 2000 N): a) displacements in the OZ axis, b) reduced stresses in the screw, c) reduced stresses in the plate

The maximum values were localized in the area of direct contact between the plate and the screw and were in the range $\sigma_{max}=2\div 160$ MPa. Increased values of reduced stresses were also observed in plate's holes – Fig. 2c. Displacement and stress distributions, obtained from the analysis of the tibia – stainless steel plate system, were presented in Fig. 2.

In another parts of the system, diverse stresses were observed, but their values did not exceed $\sigma_{max}=160$ MPa. Maximum strains were localized in areas of direct contact between individual parts of the system and were in the range $\epsilon=0.001\div 0.3\%$.

4 Results of Tibia – Ti-6Al-4V Plate System

Results of the displacements, strains and stresses analysis carried out for the tibia – Ti-6Al-4V plate system showed, that maximum reduced stresses were observed in the screws and were in the range $\sigma_{max}=0\div 68$ MPa – Fig. 3.

Also in this case, maximum stresses were localized in the area of direct contact between the plate and the screws. Stresses observed in another parts of the system were in the range $\sigma_{max}=2\div 124$ MPa. The maximum stresses were

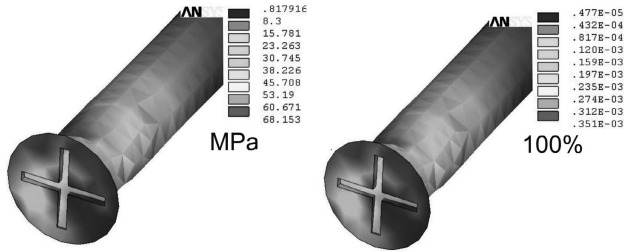


Fig. 3. Distribution of stresses and strains in the screw made of Ti-6Al-4V alloy

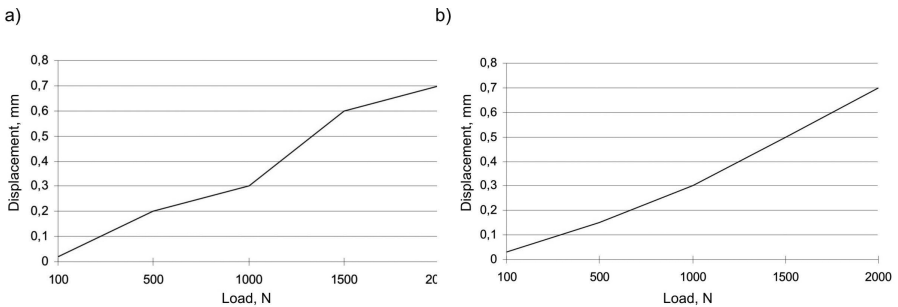


Fig. 4. Displacements in the OZ axis in a function of the applied loading: a) stainless steel, b) Ti-6Al-4V alloy

accompanied by the maximum strains. The strains did not exceed the value of $\epsilon_{max}=0.4\%$.

The important parameter influencing the effectiveness of the osteotomy is a deformability of the system. Therefore, displacements in the gap, for the applied forces from the range 100 N – 2000 N, were calculated – Fig. 4. Independently on the applied material, no significant differences in displacements (in the gap) were observed. For the plate made of stainless steel and Ti-6Al-4V alloy, displacements in the gap, for the applied maximum loading equal to 2000 N, did not exceed the value of 0.7 mm.

5 Conclusion

The aim of the work was the biomechanical analysis of the system for corrective osteotomy of tibia. Biomechanical analysis of the tibia – plate system was carried out for the implant made of two biomaterials used in bone surgery – stainless steel and Ti-6Al-4V alloy. Finite element method was applied to calculate displacements, strains and stresses. The obtained results allowed to work out biomechanical characteristics of the analyzed system. These characteristics can be a basis for selection of degree of strain hardening of the applied metallic biomaterial and optimization of the plate's geometry.

On the basis of the numerical analysis, it can be concluded that the maximum reduced stresses $\sigma_{max}=175$ MPa, obtained for the axial loading with force of 2000 N, did not exceed the yield point of the metallic material the plates were made of. Implantation of the system and loading with the maximum force of 2000 N did not cause overstressing the bone. The allowable stresses in bone, equal to 180 MPa, were not exceeded. That is important for the correct course of treatment.

References

1. Górecki, A.: Injuries of knee joint. Wydawnictwo Lekarskie PZWL, Warsaw (in Polish) (2002)
2. Ramotowski, W., Granowski, R., Bielawski, J.: Osteosynteza metodą ZESPOL. Teoria i praktyka kliniczna. PZWL, Warsaw (in Polish) (1998)
3. Granowski, Zespól – nowa metoda osteosyntezy stabilnej. Habilitation dissertation, Akademia Medyczna w Warszawie, Warsaw (in Polish) (1990)
4. Akenson, W.H., Woo, S.L.: The effects of rigidity of internal fixation plates on long bone remodeling. *Acta Orthop.* 47, 241–245 (1976)
5. Ascew, M.J., Mow, V.C., Wirth, C.R., Campell, C.J.: Analysis of the interosseous stress field to compressive plating. *J. Biomech.* 5, 203–213 (1975)
6. Carter, D.R., Vasu, R., Spengler, D.M., Dueland, R.T.: Stress fields in unplated and canine femur calculated from in vivo strain measurements. *J. Biomech.* 14, 63–70 (1981)
7. Diehl, K., Mittelmeier, H.: Biomechanische Untersuchungen zur Erklarung der Spongiosierung bei der Platten-osteosynthese. *Z. Orthop. Unfall-Chir.* 112, 235–239 (in German) (1974)

8. Strömberg, L., Dhlen, N.: Influence of rigid plate internal fixation on maximum torque capacity of long bone. *Acta Chir. Scand.* 142, 115–120 (1976)
9. Strömberg, L., Dhlen, N.: Atrophy of cortical caused by rigid internal fixation plates. *Acta Chir. Scand.* 49, 448–456 (1978)
10. Grobowski, N.T.Y.: Porównawcze badania doświadczalne zmian zachodzących w kościach długich królików po zespoleniu płytkami o różnym stopniu elastyczności. XXVII Zjazd Naukowy PTOiTr. Warsaw (in Polish) (1988)
11. Granowski, R., Ramotowski, W.: Uniwersalny płytkowy stabilizator kostny Zespol. Mater. Konf. Naukowo – Szkoleniowej Wrocławskiego PTOiTr, Karpacz (in Polish) (1985)
12. Granowski, R., Ramotowski, W., Pilawski, K.: Biomechaniczne podstawy optymalizacji osteosyntezy Zespol. XXVII Zjazd Naukowy PTOiTr Warsaw (in Polish) (1988)
13. Uthoff, H.K., Dubac, F.L.: Bone structures changes in the dog under rigid internal fixation. *Orthop. Related Research* 81, 165–170 (in Polish) (1971)
14. Veosei, V., Scharf, W., Tomiczek, H.: Anwendung von Gentamycin. PMMA – Kugeltten nach Frühinfektion unter Belastung des Osteosynthesematerials. *Unfallheilkunde* 86, 38–42 (in German) (1983)
15. Dsherov, D., Mitutzof, A.: Methoden und Mittel zur gegenwärtigen selbstspanenden Osteosynthese der Unterramknochen. *Orthop. Praxis* 6/XIV, 429–432 (1977)

Kinematic Analysis of Complex Therapeutic Movements of the Upper Limb

R. Michnik¹, J. Jurkojć¹, Z. Rak¹, A. Mężyk¹, Z. Paszenda², W. Rycerski³, J. Janota⁴, and J. Brandt⁴

¹ Silesian University of Technology, Department of Applied Mechanics, Konarskiego 18a, 44-100 Gliwice, Poland

robert.michnik@polsl.pl

² Silesian University of Technology, Institute of Engineering Materials and Biomaterials, Konarskiego 18a, 44-100 Gliwice, Poland

zbigniew.paszenda@polsl.pl

³ Upper Silesian Rehabilitation Center “Repty”, Śniadeckiego 1, 42-604 Tarnowskie Góry, Poland

wiesław.rycerski@wp.pl

⁴ Institute of Medical Technology and Equipment “ITAM”, Roosevelta 118, 41-8004 Zabrze, Poland

jacekb@itam.zabrze.pl

Summary. The paper presents the results of kinematic analysis of therapeutic movements of the upper limb, according to PNF method recommendations. Real trajectories of upper limb movements were recorded using the photogrammetric method. The measuring site consisted of a set of 8 digital cameras, two computer workstations, a set of markers, calibrating dice and light sources. On the basis of the recorded images and calculations performed with the use of specialized software, model trajectories of the analyzed movements and values of relative angular translocations and angular velocity in individual joints of the limb were defined.

1 Introduction

The number of patients requiring physical rehabilitation for loss or limitation of motion of the upper limb is growing rapidly. The cause of this phenomenon is aging of the population and the high incidence of so called civilization diseases, including disability following cerebral stroke, osteoarthritis and trauma leading to limitation of the range of motion of the limb [1, 2]. One of the methods of restoring full and permanent limb functionality is physiotherapy. For its needs, a proposition of specific physical exercises (model therapeutic motions) has been worked out, enabling the rehabilitation of single joints as well as several joints of the same limb. This model constitutes a basis of the PNF (proprioceptive neuromuscular facilitation) method, also called CPM (continuous passive motion) therapy [3, 4, 5] in the literature.

Many centers of physical rehabilitation make use of therapeutic devices, different for upper and lower limbs, based on CPM method principles. Physiological movements of the upper limb proceed normally along multiple planes (sagittal,

transverse and frontal). The majority of therapeutic devices enable only single- or two-plane motion of one or two joints [4, 5]. Thus, such a method of physical rehabilitation does not fully reproduce complex physiological movements. Considering the above, the authors started to develop a therapeutic rehabilitative device enabling the execution of multiplanar therapeutic movements. The scope of the present study included defining the real trajectories of selected (according to PNF method guidelines) complex therapeutic movements of the upper limb and defining (on their basis) the relative angular translocations and angular velocity changes in the individual joints.

2 Method

The recording of real movement trajectories was performed using the photogrammetric method. The measuring site consisted of a set of 8 digital cameras positioned to enable multiplanar recording, two computer workstations with specialized software, a set of reflective (passive) markers, calibrating cube and light sources – Fig. 1.

Smooth, multiplanar therapeutic movement of the upper limb, performed by the rehabilitator according to the PNF method, was recorded (in the primary direction – movement 1 and in the opposite direction – movement 2) – Fig. 2. The initial position of the upper limb (movement 1) was as follows:

- shoulder joint – extension approximately 0° , internal rotation approximately 45° , adduction approximately 20° ,
- elbow joint – flexion approximately 40° ,
- forearm – pronation 90° ,
- wrist – slight dorsiflexion approximately 50° .

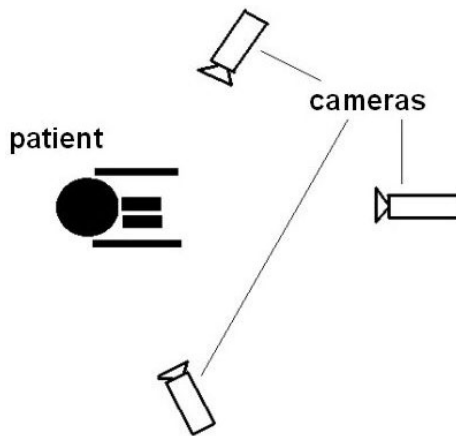


Fig. 1. Design of the measuring site for the recording of real trajectories of complex therapeutic movements

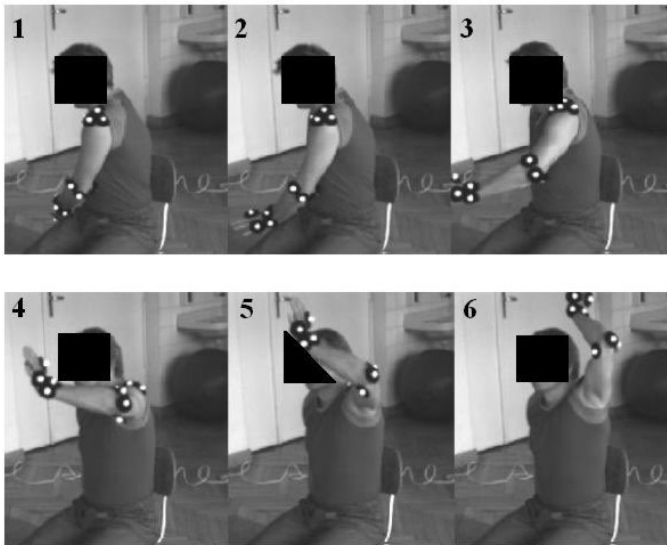


Fig. 2. Sequences of movements performed during the recording of therapeutic movement trajectories of the upper limb (movement 1)

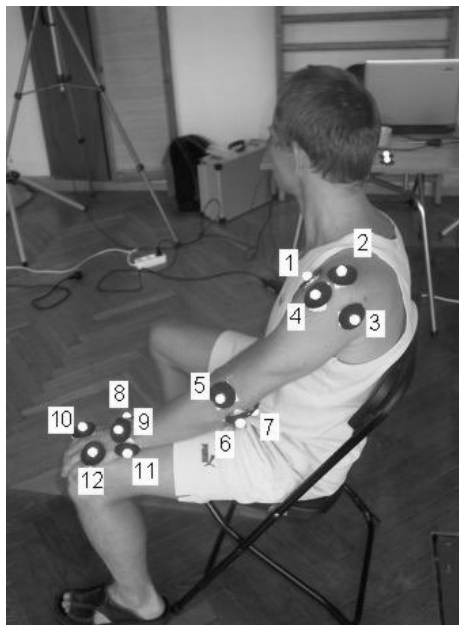


Fig. 3. Placement of reflective markers on the therapist's upper limb

The final position of the upper limb was as follows:

- shoulder joint – flexion approximately 180° , external rotation 45° , abduction approximately 160° ,
- elbow joint – flexion 90° ,
- forearm – supination 70° ,
- wrist – extended 70° .

In order to obtain simultaneous recording from all cameras, StreamPix software was used. Film processing and digitalization of the positions of individual markers placed on the upper limb was performed with APAS software – Fig. 3. The results were subsequently further processed and analyzed with the use of an unique program written in the MATLAB working environment. On that basis, the model trajectories of the therapeutic movements were defined.

3 Results

On the basis of recorded therapeutic movements (movements 1 and 2), consecutive positions of upper limb elements, as well as trajectories of individual anthropometrical points in solid orientation, were graphically depicted with the use of the photogrammetric method. This made it possible to define the course of changes in the displacement values of anthropometric points along individual axes and planes of an accepted coordinate system. Measurements results are presented in Fig. 4 – 6.

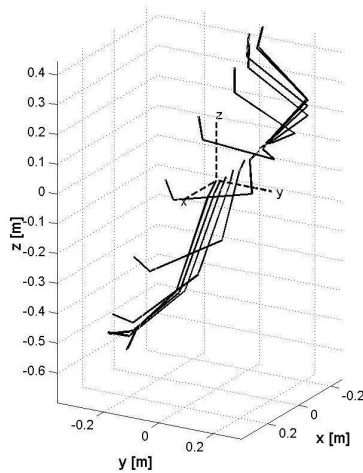


Fig. 4. Consecutive positions of upper limb elements for individual sequences of therapeutic movement (movements 1 and 2)

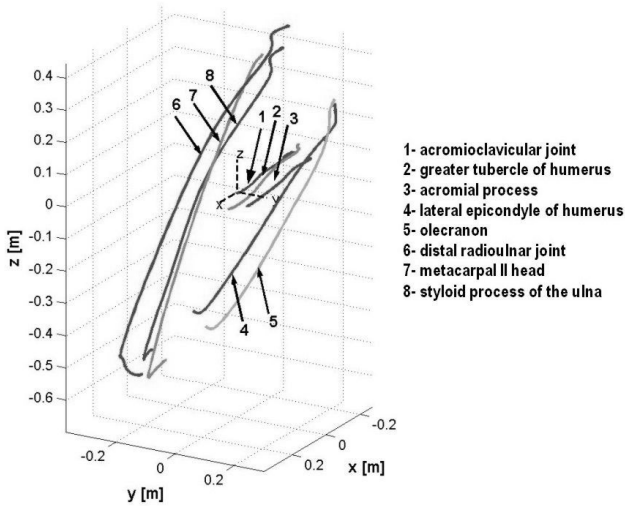


Fig. 5. Real trajectories of therapeutic movement of the upper limb (movement 1)

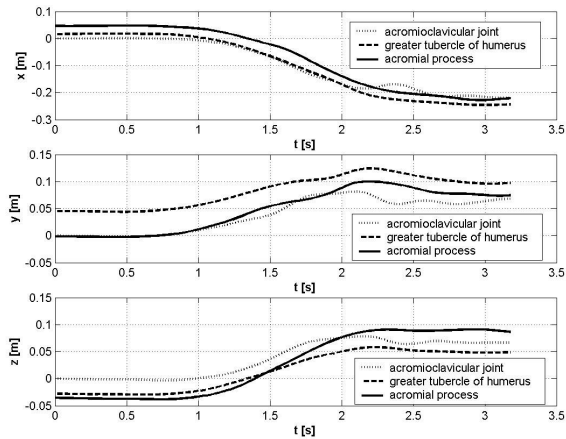


Fig. 6. Linear displacement of anthropometric points in the shoulder joint area (movement 1)

In the further part of the study, an analysis of specified, real trajectories of therapeutic movements was performed, which made it possible to define relative angular displacements and changes of angular velocity in the individual joints of the upper limb. The results are summarized in Table 1 and Fig. 7.

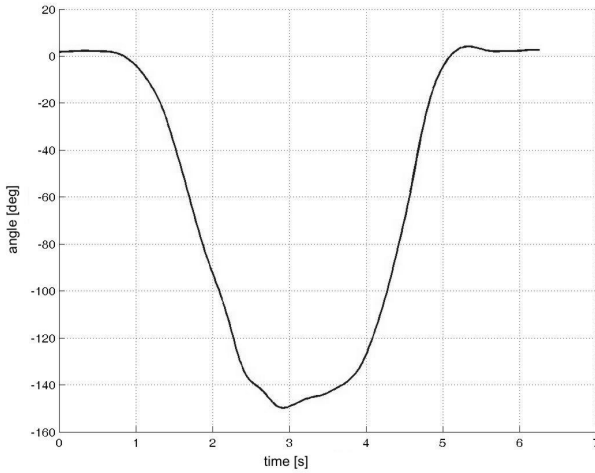


Fig. 7. Change of angle value in the shoulder joint corresponding to flexion and extension (movements 1 and 2)

Table 1. Results of kinematic analysis of therapeutic upper limb movement (movements 1 and 2)

Joint analyzed	Description		Angle range, [deg]
wrist	hand in dorsiflexion during the whole movement sequence		3÷73
elbow	flexion	limb flexed during the whole movement sequence	23÷110
	rotation	–	-110÷40
shoulder	hyperextension/ flexion	start and end of movement in adduction	4÷150
	adduction/ abduction	start and end of movement in adduction	-12÷140
	external / internal rotation	–	5÷100

On the basis of measurements and calculations performed, model courses of therapeutic movements were worked out, according to PNF method guidelines. They were defined by establishing the mean values of the kinematic parameters in the recorded movement sequences – Fig. 8, 9.

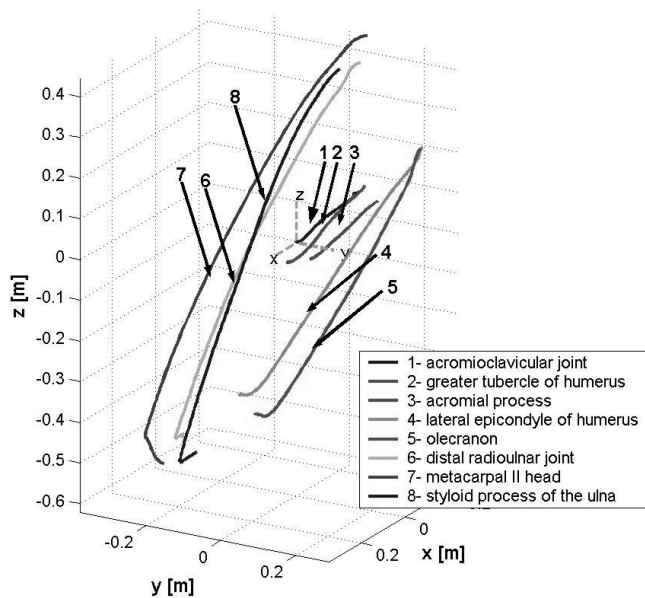


Fig. 8. Model trajectories of therapeutic movement of the upper limb (movement 1)

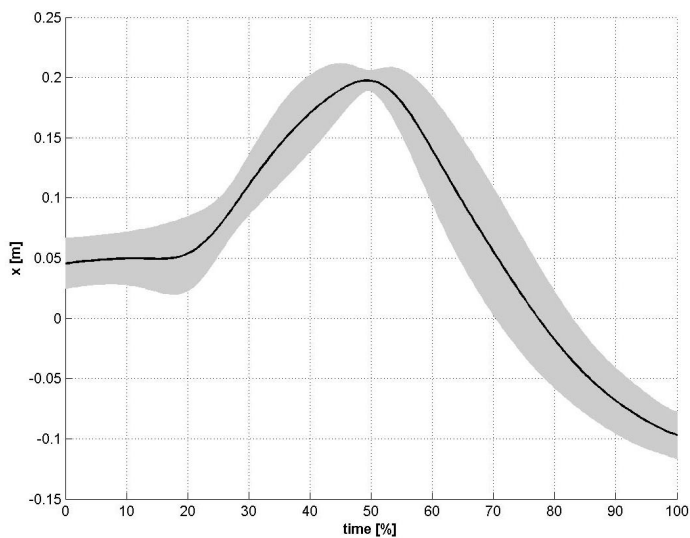


Fig. 9. Averaged course of point 6 linear displacement along the X-axis (movement 1)

4 Conclusions

The photogrammetric method used in this study made it possible to define model trajectories and the kinematic parameters of complex, multiplanar therapeutic movements of the upper limb. From among the therapeutic movements recommended by the PNF method, only the complex, multiplanar movements shown in Fig. 2 were analyzed in the present study.

The model trajectories of therapeutic movements and the results of the kinematic analysis obtained in this study will serve as a basis for the development of a device for upper limb physical rehabilitation, capable of reproducing the complex movement sequence.

The values of linear displacement observed in the shoulder joint deserve special attention – Table 1, Fig. 7 and 8. In complex movements they reach values which should be taken into account during the design of the rehabilitative device's kinematic parameters as well as its control.

Acknowledgement. The study was supported by research grant no. R13 027 02 of the Ministry of Science and Higher Education.

References

1. Marciniak, J.: Hospital and rehabilitation equipment. Printing House of the Silesian University of Technology, Gliwice (in Polish) (2005)
2. Kiwerski, J.: Medical rehabilitation. PZWL, Warsaw (2005)
3. Mirek, E., Chwala, W., et al.: Proprioceptive neuromuscular facilitation method of therapeutic rehabilitation in the treatment of patients with Parkinson disease. *Neurol. Neurochir. Pol.* 37, 89–102 (2003)
4. Blauth, W.: CPM therapy with motorized exercise devices. Urban&Vogel, München (1992)
5. Mavroidis, C., Nikitzuk, J., Weinberg, B., et al.: Smart portable rehabilitation devices. *Journal of NeuroEngineering and Rehabilitation* 2 (2005)
6. Michnik, R., Jurkojc, J., Jureczko, P. Metody modelowania ruchu kończyn dolnych człowieka w trakcie chodu. In: Proceedings of the 3rd Scientific Conference „Materials, Mechanical and Manufacturing Engineering” M³E’2005, Gliwice – Wisła, pp. 809-818 (2005)
7. Pieniazek, M., Chwala, W., Szczechowicz, J., Pelczar, M.: Upper limb joint mobility ranges during activities of daily living determined by three-dimensional motion analysis. *Ortopedia Traumatologia Rehabilitacja* 4, 413–422 (2007)

Influence of Model Discretization Density in FEM Numerical Analysis on the Determined Stress Level in Bone Surrounding Dental Implants

Jarosław Żmudzki¹, Witold Walke², and Wiesław Chladek¹

¹ Department of Technological Processes Modelling and Medical Engineering, Silesian University of Technology, 40-019 Katowice, Poland

{jaroslaw.zmudzki,wieslaw.chladek}@polsl.pl

² Institute of Engineering Materials and Biomaterials, Silesian University of Technology, 44-100 Gliwice, Poland

witold.walke@polsl.pl

Summary. Influence of mesh density on the results of FEM model analysis of mechanical biocompatibility of dental implants has not been presented yet. Taking advantage of the Ansys v.11 software, carried out was an analysis in the linear elastic range level of stresses in bone tissue surrounding standard osseointegrated implant of complete denture, with a decreasing size of elements (tetragonal type SOLID 187) adjacent to cortical bone/implant interface, respectively 0.5, 0.3 and 0.1. Equivalent Huber-Mises' stress value in the zone that is exposed to effort the most, and is located close to the edge of implant insertion into the cortex bone significantly increases along with mesh density from app. 60 MPa to 120 MPa for opposite models, because of the lack of convergence of stresses at this singularity point. Increase of mesh density leads to an overestimation of loading stresses values and furthermore to an unjustified increase of pillars' diameter. At the other hand, too large elements might lead, through an underestimation of loading stress level, to overloading atrophy of bone tissue or to implant loss.

1 Introduction

Assessment of mechanical loading of implants and surrounding tissues is one of the tasks of biomaterials engineering. The basic problem is not only the selection of solutions for mechanical properties as far as their strength or wear are concerned, but at the first place, paying attention to the optimal loads distribution in tissues around implants. Researches are especially focused on evaluation of bone tissue loadings, affected by tempestuous remodeling processes around intra-osseous implants that result from loads redistribution in respect of the natural state.

Typical example here are dental implants, for which 90% of clinical failures based on implant loss result from biomechanical reasons, and only 10% are caused by biological phenomena, the so called "periimplantitis" [6].

It is most effective to carry out the evaluation of structure's strength and its influence on the adjacent tissue at the conceptual stage by means of a FEM

analysis. That method makes it possible, already at the early design stage, to eliminate solutions favoring atrophic processes caused by tissue over- or under-loading. Known are the examples of foreseeing far future clinical effects on the basis of a bone tissue density pattern around implants while taking into account the remodeling or contact phenomena, although the first step of bone loading FEM evaluation should always be the linear analysis. Results of this analysis should be taken as a point of reference for further analyses that would have a higher degree of projection of the complex real system, that are very sensitive to experiment conditions, which influences the remarkable spread of results and creates difficulty with their interpretation [2, 8].

Hence, in the search for real stress values, the influence of model division on finite elements, on results spread, mainly in the foreseen stress concentration zones should be checked. These zones are assumed to be the criteria defining areas, which decide whether or not given constructional solution is appropriate. Skipping this analysis makes it impossible to evaluate what is the relation between a discretized model and the exact solution, let alone the real state. Appropriate selection of discretization parameters is not possible in a general way, as it mainly depends on specific geometric and material characteristics of the examined system, as well as on the assumed computational technique. Therefore, the only one way is to check the given issue “in practice” with a given FEM software, where discretization density is differentiated on purpose.

Biomechanical analyses of denture implants are carried out quite often, however most of them only state in their parts related to methodology the name of the used software, rarely the type of elements or the total number of elements. Sometimes, given is, in a general way, the confirmation of carried out analysis of the number of elements on the displacement convergency. Mesh parameters are not given and they are usually very difficult to be recognized on the basis of the sometimes shown models. On such a basis it is not possible to define to which extent the analyses results from various sources are comparable with each other, especially in respect to the criteria defining tissue areas in stresses concentration zones.

Hence, in this research special attention has been paid to determining the influence of basic FEM model experiment conditions connected with mesh density on the obtained results in the commonly used linear models. The scope of that research has covered stresses levels in bone tissue surrounding a standard denture implant, using an ANSYS v.11 software.

2 Materials and Methods

FEM model researches regarding loading of bone tissue surrounding the implant of complete lower denture have been carried out by means of ANSYS v.11 software in the linear-elastic scope. Computer parameters were as follows: CORE Quad Intel®Xeon®CPU E5335@, 2GHz 16382GB FB DDRAM 2, FSB 1330 MHz; Win Vista 64 bit. Because of the software limits, only two cores of the processor have been used. For the purposes of the carried out analysis bone

geometry has been simplified to a cylindrical shape surrounding the anchorage area.

Model cross-section presenting the assumed system of layers is shown on Fig. 1. The pillar was loaded with a force of 150N [11] applied at the angle of 45 degrees to implant's axis. Assumed complete adherence of the implant to the bone is reflected by the state after osseous-integration phase. Model has been fixed to the bottom and lateral surface of the bone. Division into finite elements has been carried out at various mesh density levels, close to the edge of pillar insertion into the cortex bone. Control over the size of the elements has been achieved by means of a primary division of the interface cortical bone/implant with a pre-set value of 0.1, 0.3 and 0.5. In this way prepared have been models that have various sizes of elements: 281 796, 23 029 and 13 500. 10-Node Tetrahedral have been used (AnSys Solid 187) [1].

Assumed were advantageous osseous foundation conditions: cortex bone thickness 2.0 mm; cortex bone's Young's modulus of elasticity 16 000 MPa, Poisson's coefficient $\nu=0.3$, and for spongy bone: $E=600$ MPa; $\nu=0.4$, whereas $E=110\,000$ MPa and $\nu=0.3$ have been assumed as average characteristics for titanium alloys, of which pillars are to be made.

For the purposes of error estimation for the calculated stresses values, an option "energy error per element" available for linear analyses has been used (SERR ANSYS' command) according to the method shown in paperwork [16]. By summarizing all element error energies e_i , the global energy error in the model e , can be determined. This can be normalized against the total energy ($u + e$),

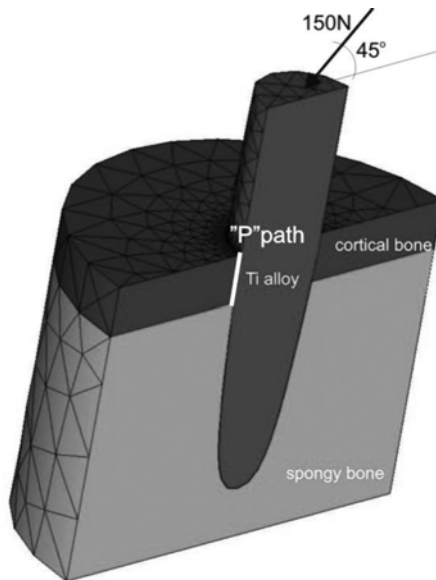


Fig. 1. Conditions of FEM model analysis

where u is the strain energy, and expressed as a percent error in energy norm, E . The percentage error in energy norm E (SEPC ANSYS' command) is indicated as a good overall global estimate of the discretization or mesh accuracy [5]. The maximum absolute value of nodal stress variation of any stress component for any node of an element (SDSG ANSYS' command) was also taken into account.

3 Results

Comparisons of mesh density influence on the analyses results have been carried out for equivalent stresses (acc. to the Huber-Mises hypothesis).

On Fig. 2a the pattern of equivalent stresses in cortex bone has been shown. The highest stresses level is discovered on the surface, close to the edge in the plane of a highest bending caused by the horizontal component of the loading force. Differences in stresses values between models having opposite meshes are huge. Stresses discovered in case of a low mesh density (Model 0.5) are app. 60 MPa, whereas for a high density mesh (Model 0.1) they reach values above 120 MPa. Equivalent stresses in the implant have not exceeded 205 MPa.

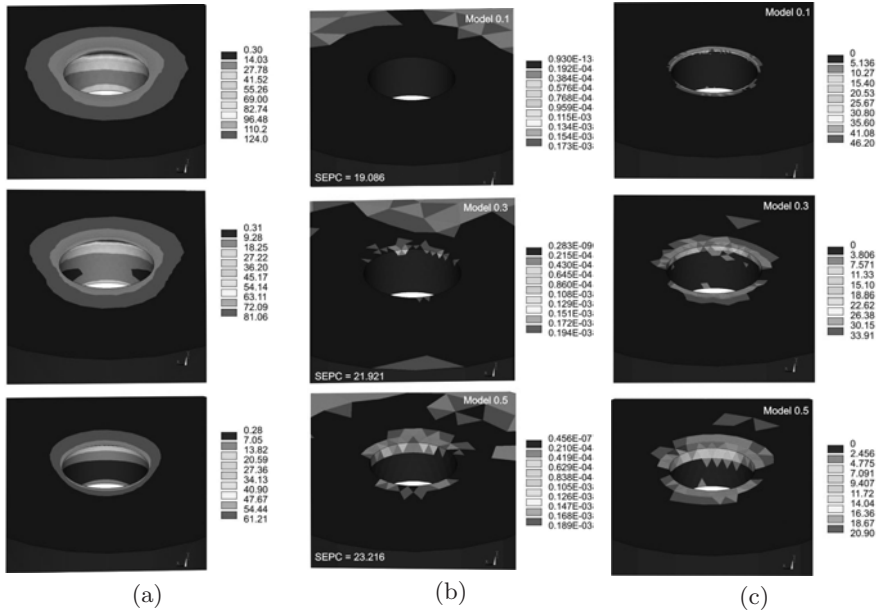


Fig. 2. (a) Equivalent stresses (Huber-Mises') in cortical bone for all the tested meshes density, MPa; (b) Energy error per element for all the tested meshes, as well as an amount of structural percentage error in energy norm: SEPC; (c) The maximum absolute value of nodal stress variation of any stress component for any node of an element for all the tested meshes, MPa

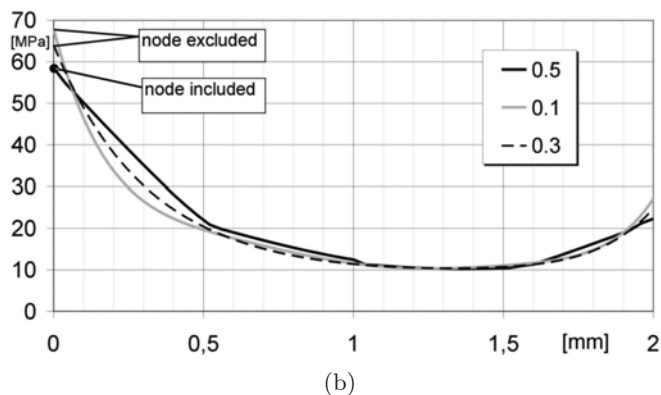
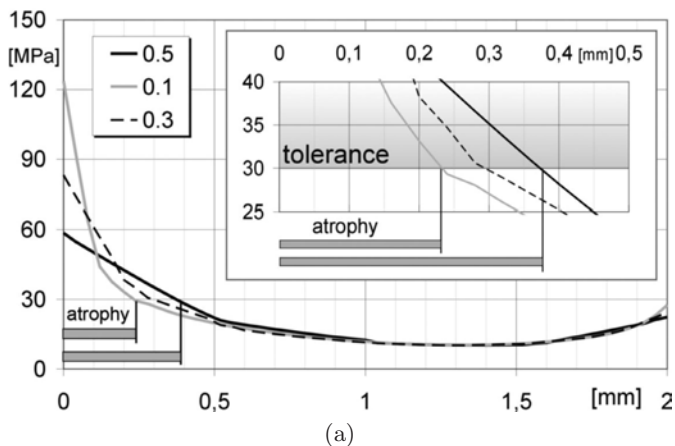


Fig. 3. (a) Profiles of equivalent stresses made on the basis of node values along “P” for all the tested meshes; (b) Profiles of equivalent stresses achieved on basis polynomial approximating nodal values after excluding the maximum value close to the edge (singularity point) for meshes 0.1 and 0.3

On Fig. 2b there is shown an energy error (SERR) per element and the value of structural percentage error in energy norm (SEPC) for all of the tested meshes.

On Fig. 2c there is presented the maximum absolute value of nodal stress variation of any stress component for any node of an element (SDSG) for all the tested meshes.

On Fig. 3a for all the tested meshes there has been presented a profile of stresses in cortex bone, made on the basis of node values along “P” path shown on Fig. 1.

Then in Fig. 3b, in model, the peak values were excluded in the node closest to the edge (but not in case of coarse mesh model 0.5). The maximum value close to the edge has been regained after introduction of the polynomial approximating all the remaining values.

During FEM results analysis forgotten should not be the differences between the model and the real system, that result from many, necessary in model researches, simplifying assumptions. Biomechanical models are complex, as far as their geometry and materials are concerned. Tissue geometry is usually achieved by means of spatial technique of reconstruction of 2-dimensional tomography scans. Prosthetic constructions are built directly by means of CAD systems [7, 14]. In case of tomography scans reconstruction, there occurs a problem of a proper mapping of bone outline and the boundary between cortical and cancellous tissue or a problem of a precise determining of bone density. Such inaccuracies directly affect stresses values computed by means of a FEM method. It has not been, however the object of these considerations. This research is focused on influence on stresses values of discretization method, having the geometry and elastic properties already established. Therefore, justified is here also the assumption for the model of a simplified shape of the bone surrounding the anchorage zone and the shape of implant without thread or abutment internal components. Model discretization has been performed by means of the used tetrahedral type of finite elements, as because of the complex shapes of prosthetic structures and especially those of bones, the use Hexahedral type of elements is difficult [14].

A success of a prosthetic treatment depends directly on the risk of implant loss resulting from overloading, bone tissue atrophy or fatigue implant failure. Pillar bending loads are in this respect very unfavorable [4, 6]. They increase the risk of pillar exposure off the bone and even a risk of a fatigue fracture. In the analyzed model, level of stresses, only in the pillar of the dimension of 4 mm is significantly lower than the fatigue strength of titanium alloys. It has to be pointed out, however, that the results are limited to a filled single-part implant without abutments, assuming the 5 mm long lateral forces lever arm. Nevertheless, it is not the implant construction that the paper is focused on, but the bone loading possibilities.

The level of equivalent stresses for cortex tissue in the area of pillars' entrance into the bone is dangerously high. In spite of the fact that the values are located below the average cortex tissue strength, for the purposes of increasing the tolerances of cycling loads even a lot more lower values should not be exceeded in case of shear stresses of 30–35 MPa [15].

Full adherence, assumed for the tested model creates a situation of a symmetric stresses pattern differentiated by the minus- or plus-sign. Stresses results from the pillar compression and tension at both sides of the bent pillar evoked by pillar bending. In the phase before osseointegration stresses will achieve much higher level, as a portion of loadings at the side of tension cannot be borne, and the whole loading has to be borne at the compression side.

The key element of the atrophic processes simulation is the assignment of the decreased density and modulus of elasticity in overloaded zones during the sequential calculating steps [3, 15]. In the analyzed case, in the first place the atrophic process affects the most overloaded areas adjacent to the edge, only in the highest bending plane [15]. As a result of atrophic changes the geometry of the bones around the edge and the manner of loads bearing get completely

changed. Implant does not press the edge anymore in the plane of the highest bending, because of the depression formed in that area, but it starts pulling the bone on lateral edges, which were earlier not exposed to such effort and therefore have not atrophied. If, however stresses do not exceed tolerances, then the process can be stopped. In an opposite situation this process proceeds, leading at its further stages to a pacing exposure of the implant off the bone [15]. Model, analyzed in this work refers only to the first calculation step, nevertheless, the discretization parameters might have their reference in the next steps.

Dangerous stresses values registered at the edge in the tested model have significantly increased along with the increase of mesh density; from app. 60 to more than 120 MPa. On the other hand, the area of the overloaded bone gets narrower with a more dense mesh.

The problem of the influence of division into finite elements on results is commonly known, however the basic rule is to carry out comparative models researches always with identical mesh parameters. During evaluation of the tissue loading, apart from comparative targets of various solutions, searched are also the absolute stresses values. The general rule says about an optimal mesh size, because although while increasing mesh's density decreased is also the method's error that results from the unknown values between the nodes (too coarse discretization division), increased is the influence of computational algorithms and the round-off error.

In the presented analysis, increasing mesh density results in an increased stresses value at nodes adjacent to the edge, which constitutes in the function of stresses a special singular point because of the geometric and material notch [12]. Displacements of the tested meshes have proved to be converged as the differences have not exceed 1-3%. Differences between nodal values on the remaining length of the adherence zone are insignificant for the examined meshes. The SEPC value, determined on the basis of SERR, significantly decreases along with increase of mesh density (Fig. 2b). Although, this parameter shows a better discretization and precision of stresses function in the whole of the bone, stresses peaks in the last of the nodes that is close to implant, in case of a dense mesh are characterized by a higher uncertainty (Fig. 2c).

On the basis of the hitherto deliberations proved by the analysis of calculations errors it is still very difficult to determine whether the stress values at the level of 60 or 120 MPa can be assessed as the ones reflecting the real loadings. On the basis of the work [13] it can be assumed that a precise solution is impossible to be achieved, because the stresses around singularity points do not reach the convergency. The only one possibility is the labor-consuming and expensive experimental verification [10], which is however unnecessary at the conceptual stage. Hence, in a numerical analysis it is optimal to assume that the peak value is of no relevance and that the whole bone area, in which occurring stresses exceed tolerances limits undergo the atrophic changes. In case of a search for an iso-surface that separates the overloaded and the not-overloaded tissue areas, a coarse mesh results in an increase of that area (Fig. 3a). Decrease of the elements adjacent to the edge to app. 0.3 increases precision of determining the iso-surface location.

One can, however imagine a situation, in which stresses values reach tolerance criterion only at a node in an uncertain singularity point. Should, in such situation such a value be excluded and should it be assumed that an atrophic process will not take place here? Some possibility of getting more precise stresses values is the use of linear hexahedral elements [13], although such values apart from being less sensitive to refinement of mesh at singularity points, are still uncertain and generating such meshes in case of complex geometries might be very time-consuming. On the other hand, stresses values located close to singularity points, in case of a dense mesh could be precisely estimated. As the values in the singularity point should not be taken into account [3, 9, 13], the authors of this research state that such values should be completely excluded. Then, having enough number of remaining nodal values, the extreme value can be again extrapolated as shown on Fig. 3a. Nevertheless, in case of a coarse mesh, an extreme value obtained in the above mentioned way is affected by a remarkable uncertainty, as the insignificant differences of any of the available 5 nodal values result in a notable change of the OY axis crossing point. Hence, for the 0.5 model the node closest to the edge has not been removed. At the same time, it can be observed that the extreme value for the model 0.3 and 0.1, after exclusion of the extreme node, comes closer to the nodal value of the coarse mesh. In case of a coarse mesh there is an underestimation reaching app. 10 MPa. It denotes that, for the analyzed model, assumption of the elements slightly smaller than 0.5 only close in the area to the edge would make it possible to achieve stresses values close to the ones obtained in the solution with a dense mesh, after a labor-consuming processing without any possibility of using a FEM software.

Therefore, according to the conclusion drawn in research [13] it has been proved that an excessive increase of mesh density around singularity points leads to a notable spread of stresses values along with a remarkable increase of the computational cost, at the same time making the increase of mesh density ineffective. As it has been proved in this work, a more precise determining of iso-surface covering overloaded bone areas is possible, although an only insignificant overestimation of that area, for a coarse mesh, gives as a result the design solution that is far more safe for the bone.

4 Conclusion

In evaluation of bone tissue loadings, stresses peaks occurring around singularity points, that are normally exposed by strength analysis software as criteria values, should not be taken into account. The most effective evaluation criterion is to give up the stresses comparisons and to determine the amount of the overloaded bone, while defining a relatively coarse mesh.

In case of a too dense mesh, achieved stresses values are overestimated, which leads to unnecessary increase of implants diameter. On the other hand, in case of a too coarse mesh, due to the underestimated values, there is a risk of causing an overloading atrophy of cortex bone or formation of undesirable connected tissue.

If necessary, the extreme values' convergency can be searched by means of increased mesh density, excluding the nodes adjacent to singularity points, and then extrapolating stresses function along the path that starts far beyond the range of singularity points' interactions.

References

1. Ansys element reference, part I Element library, solid 187
2. Chabanas, M., Payan, Y., Marecaux, C., Swider, P., Boutault, F.: Comparison of linear and non-linear soft tissue models with post-operative ct scan in maxillofacial surgery. In: Cotin, S., Metaxas, D.N. (eds.) ISMS 2004. LNCS, vol. 3078, pp. 19–27. Springer, Heidelberg (2004)
3. Chen, G., Pettet, G., Pearcy, M., Mcelwain, D.L.S.: Comparison of two numerical approaches for bone remodeling. *Medical engineering & physics* 29, 134–139 (2007)
4. Chladek, W., Majewski, S., Żmudzki, J., Krukowska, J.: The mechanical condition of the functionality of implantoprosthesis construction - model investigation. *Implantoprotetyka* 2, 3–10 (2003)
5. Error approximation technique for displacement-based problems, Release 10.0 doc. For ansys, ch. 19.7.1
6. Esposito, M., Hirsch, J.-M., Lekholm, U., Thomsen, P.: Failure patterns of four osseointegrated oral implant systems. *J. Mater. Sci. Mater. Med.* 8, 843–847 (1997)
7. Jaecques, S.V.N., van Oosterwyck, H., Muraru, L., van Cleynenbreugel, T., de Smet, E., Wevers, M., Naert, I., van der Sloten.: Individualised, micro ct-based finite element modelling as a tool for biomechanical analysis related to tissue engineering of bone. *Biomaterials* 25, 1683–1696 (2004)
8. Lee, W.C.C., Zhang, M., Jia, X., Cheung, J.T.M.: Finite element modeling of the contact interface between trans-tibial residual limb and prosthetic socket. *Med. Eng. Phys.* 26, 655–662 (2004)
9. Mańko, Z., Duchaczek, A.: Development of fatigue cracks in erection junctions of low water bridges' beams, *Scientific booklets, wsow łąd.* 2(136), 26–38 (2002)
10. Ramos, A., Simoes, J.A.: Tetrahedral versus hexahedral finite elements in numerical modelling of the proximal femur. Technical note, *Medical engineering & physics* 28, 916–924 (2006)
11. Sahin, S., Cehreli, M.C., Yalcin, E.: The influence of functional forces on the biomechanics of implant-supported prostheses - a review. *J. Dent.* 30, 271–282 (2002)
12. Stolk, J., Verdonschot, N., Huiskes, R.: Sensitivity of failure criteria of cemented total hip replacements to finite element mesh density. In: 11th Conference of the ESB, Toulouse, France, July 8–11 (1998)
13. Stolk, J., Verdonschot, N., Huiskes, R.: Management of stress fields around singular points in a finite element analysis. In: Middleton, J., Jones, M.L., Shrive, N.G., Pande, G.N. (eds.) *Computer methods in biomechanics and biomedical engineering*, pp. 57–62. Gordon and Breach Science Publisher, London (2001)
14. Teo, J.C.M., Chui, C.K., Wang, Z.L., Ong, S.H., Yan, C.H., Wang, S.C., Wong, H.K., Eeoh, S.H.: Heterogeneous meshing and biomechanical modeling of human spine, *Communication. Medical engineering & physics* 29, 277–290 (2007)
15. Van Oosterwyck, H., van der Sloten, J., Puers, R., Naert, I.: Finite element studies on the role of mechanical loading in bone response around oral implants. *Meccan* 37, 441–451 (2002)
16. Zienkiewicz, O.C., Zhu, J.Z.: A simple error estimator and adaptive procedure for practical engineering analysis. *Int. J. Num. Method. Eng.* 24, 337–357 (1987)

Computer Simulations of Electric Properties of Organic and Non-organic Compounds

P. Janik¹, M.A. Janik², and Z. Wrobel¹

¹ Instytute of Informatics, University of Silesia, Bedzinska 39, 41-200 Sosnowiec, Poland

pjanik@us.edu.pl

² Department of Statistics, Medical University of Silesia, Ostrogorska 30, 41-200 Sosnowiec, Poland

Summary. In this contribution, we analyse the dielectric response of a simple RC model composed of single RC circuits (relaxors) with the exponential response. The relaxation time of a single relaxor is defined by the product of resistance and capacitance: $\tau = RC$. We present two hypothetical algorithms of generating gaussian input function, one with fluctuation of a density decay, and the other one with normalized density. Both of them enable to obtain a different impedance response of the system. An algorithm of input function creation is crucial to generating the response of the presented model.

1 Introduction

The measurements and the analysis of impedance are of great importance in such research methods as, e.g.: spectroscopy [1] or impedance tomography [2]. The measurements of impedance are a non-intrusive method of studying both organic and non-organic samples. Their additional advantage are a short measurement time and a relatively simple analysis of measurement results [3]. For this reason, they can be "attractive" in biomedical applications. However, this method may also prove defective. For example in the case of a wrong choice of measurement electrodes, electric parameters of measurement apparatus, or a surface of an electrode-sample contact etc. Although spectroscopy or impedance tomography give fast and good measurement effects, they are not easy methods as they require some measurement experience and basic knowledge of objects which are studied. One of the methods whose aim is to improve the measurements techniques and the search for the genesis of some phenomena which take place in a given sample is modelling and simulation of real systems. Modelling of impedance objects has been done for many years now in different domains of science such as physics, electrical engineering, electronics, metrology, biomedical engineering. Although literature abounds in electrical models [4]-[13], a universal system which could reflect complex processes taking part in real systems is yet to be invented. To the best of the authors' knowledge, a model which takes into consideration both processes in a studied sample and apparatus artefacts (such

as electrode effects, an existence of double-layer etc.) does not exist. In this paper an RC model, which enables to take into consideration the afore-mentioned phenomena, is presented. One can divide a simulated object into areas which have a different relaxation time τ . Real systems, biological samples in particular, are not a set of identical relaxors but it is possible to distinguish in them smaller and bigger particle clusters which relax in the same way and thus, are characterized by a certain distribution.

1.1 Description of RC Model

A circuit diagram, named "lattice-like", built with simple RC elements (relaxors), presented in Fig. 1, were analyzed. Relaxors were connected in a series and limited by two nodes which reflect the plates of the capacitor. Relaxation times of particular relaxors are defined by the product of resistance and capacitance. The the resistance is a dissipative element, whereas the capacitance describes an inertia of the system (an amplitude of the real part of dielectric permittivity). The relaxation times can be distributed by changing the R and/or C value. In the presented model the value of particular capacitors was assumed to be the same and equal to 1nF, whereas, the value of resistors changed according with Gaussian distribution (in logarithm scale). Additionally, the interactions between particular RC sections, realized by resistors R_S , were taken into consideration. Moreover, in order to eliminate the limitation of the used simulator, resistors R_{sym} had to be inserted.

The simulation process was carried out using PSPICE software.

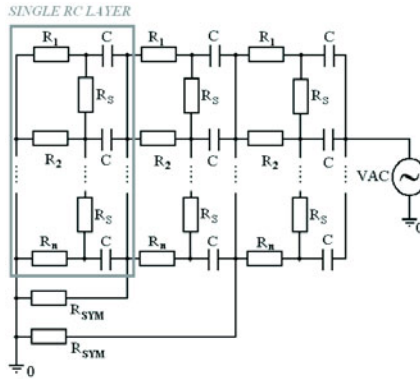


Fig. 1. Diagram of the RC model

1.2 Distribution of Relaxation Times

In the presented RC model the Gaussian distribution was used, according with the following relation:

$$n_1 = \frac{A}{\omega\sqrt{\frac{\pi}{2}}} \exp \left[\frac{-2(\tau_i - \tau_A)^2}{\omega^2} \right] \tag{1}$$

where n_1 - a number of RC elements with the same relaxation times τ_i , A - an area under the appropriate plot, τ_A - a mean value of all relaxation times used in appropriate plot, ω - the half width of the Gaussian curve. The Gaussian distribution in logarithmic scale was used, due to the limitation of this distribution in the linear scale (the left wing of the Gaussian function reaches negative (non-physical) values).

2 Results and Discussion

In the simulating model the Gaussian distribution of relaxation times in logarithmic scale was chosen. The half width of the Gaussian curve (ω eq. (1)) was about four decades (Fig. 2a, Fig. 2b) in order to obtain stretched relaxation peak (Fig. 2d). Changing ω towards higher values made the response curve more stretched [14]. The simulating circuit is an electronic summator of relaxation times of its all single relaxors. The chosen distribution determines a number of RC sections with appropriate relaxation time ($\tau = 1/(2\pi f_p)$) where f_p - a peak frequency). In order to model the responses of complex relaxation processes, one should not be limited to the choice of a suitable distribution function. It is vital to take density decay of points per decade into consideration. Fig. 2a and Fig. 2b show two identical decays, which differ only in terms of density decay.

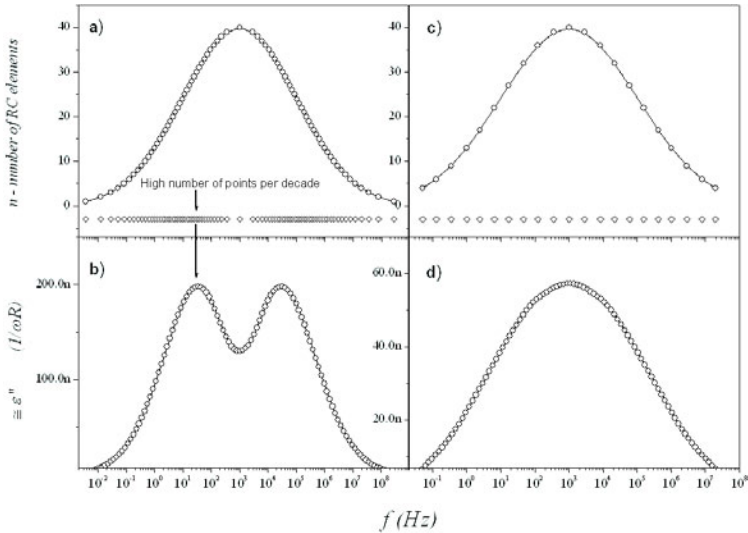


Fig. 2. The response of the model with gaussian distribution of relaxation times with fluctuation of a density decay a), b) and with normalized density c), d)

This is visible in a projection of points on a straight line under an appropriate figure. In Fig. 2a it is possible to notice areas with an increased density of points which on weight summing realized by the circuit generate two local peaks (Fig. 2c). A completely different situation takes place when a decay is normalized (e.g. a constant number of points per decade) (Fig. 2b), which enables to generate responses consistent with the shape of a given distribution of relaxation times (Fig. 2d). There are many groups of materials which have a similar character of impedance curves, but they differ considerably in terms of, for example, position of peak maximum or its amplitude. The presented model enables the change of peak amplitude in proportion to capacitance value C and the change of peak maximum frequency with the τ_A value (Fig. 3). Though the models which have been presented in literature so far are electrically complex and can generate complex impedance response [10]-[13], they are not capable of modelling a wider group of materials. The model presented in this contribution has a constant electrical configuration of a circuit. Still, it is possible to generate any number of impedance responses owing to the application of algorithm of relaxors' distribution. Alike in nature, many quantities are characterized by a certain decay of values (distribution). Owing to the afore-mentioned characteristics, the presented model is more universal and the obtained results are easier to interpret. The model also takes into consideration coupling among particular RC sections (resistor R_S in Fig. 1), which expands its possibilities to a considerable extent. Owing to this it enables to generate the characteristics of complex systems (multifractional) [15]. However, this aspect has been omitted in the contribution.

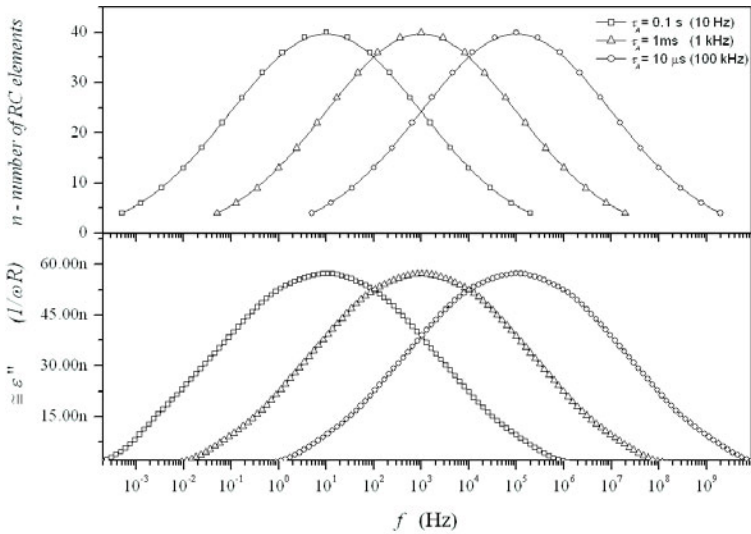


Fig. 3. The migration of the frequency peak depending on chosen relaxation time

3 Concluding Remarks

The presented model enables to obtain any number of impedance response according to the given distribution of relaxation times. In practice it means that if we want to generate a response curve of some shape which is similar to the spectrum of real system, it is enough to choose algorithm distribution (function and its density decay) which best reflects this spectrum. This model has only described relaxation phenomena so far. However, it can be easily applied to resonance phenomena as well.

References

1. Kremers, F., Schonhals, A.: Broadband dielectric spectroscopy. Springer, Berlin (2003)
2. Holder David, S.: Electrical Impedance Tomography: Methods, History And Applications. Inst. of Physics Pub. Inc. (2005)
3. Dyszkiewicz, A., Janik, P., Janik, M.: *Physiol.*, vol. 2, pp. 1345–1359 (2006)
4. von Hippel, A.: Dielectrics and waves. John Wiley and Sons, Inc., New York (1959)
5. Macdonald, J.R.: *J. Appl. Phys.* 58, 1955–1970 (1985)
6. Macdonald, J.R.: *J. Appl. Phys.* 62, R51–R62 (1987)
7. Beekmans, N.M., Hayne, L.: *Electrochim. Acta.* 21, 303–310 (1976)
8. Bruce, P.G., West, A.R.: *Electrochem. Soc.* 130, 662–669 (1983)
9. Perez, E., Wolfe, J.: *Eur Biophys. J.* 16, 23–29 (1988)
10. Han, D.G., Choi, G.M.: *Journal of Electroceramics* 21, 57–66 (1998)
11. Dissado, L.A., Hill, R.M.: *Phys. Rev. B* 37, 3434–3439 (1988)
12. Dyre, J.C.: *Phys. Rev. B* 48, 12511–12526 (1993)
13. Vainas, B., Almond, D.P., Luo, J., Stevens, R.: *Solid State Ion.* 126, 65–80 (1999)
14. Janik, P., Paluch, M., Tomawski, L., Ziolo, J.: *Eur. J. Phys.* 21, 233–237 (2000)
15. Janik, P., Janik, M., Tomawski, L., Ziolo, J., Paluch, M.: *J. Non-Cryst. Solids* 353, 3932–3935 (2007)

Author Index

- Augustyniak, Piotr 3, 243
- Badura, Paweł 192
- Basiaga, Marcin 529
- Błaszczński, Jerzy 497
- Bojar, Konrad 129
- Boninski, Piotr 113
- Brandt, J. 551
- Brzoza, Piotr 513
- Buchalski, Zbigniew 461
- Burduk, Robert 476
- Chachulski, Damian 335
- Charytanowicz, Małgorzata 437
- Chladek, Wiesław 559
- Choraś, Ryszard S. 121
- Cichosz, Jacek 217
- Cieplak, Jerzy 521, 545
- Czabański, Robert 283, 299
- Czarnecka, Anna 157
- Dec, Stanisław 315
- Doros, Marek 315
- Dulewicz, A. 403
- Dyzkiewicz, Andrzej 335
- Glinkowski, Wojciech 28
- Goszczyńska, Hanna 315
- Graczyk, S. 325
- Grzanka, A. 325
- Grzymała-Busse, Jerzy W. 225
- Henzel, Norbert 283
- Hippe, Zdzisław S. 225
- Horoba, Krzysztof 291, 299, 307
- Hudyma, Elżbieta 157
- Iwanowski, Marcin 365
- Janik, M.A. 568
- Janik, P. 568
- Janota, J. 551
- Jaszczak, P. 403
- Jezewski, Janusz 291, 299, 307
- Jezewski, Michał 299
- Jozwiak, Rafał 208
- Jozwik, Adam 488
- Jurkojć, J. 551
- Kaczmarek, Marcin 521, 537, 545
- Kajzer, Wojciech 537
- Kartaszynski, Rafał Henryk 184
- Kawa, Jacek 149
- Kiel, Marta 529
- Kolebska, Krystyna 315
- Koprowski, Robert 232
- Korzynska, Anna 345, 365
- Kotas, Marian 251
- Kowalczyk, Leszek 315
- Krauze, Anita 537
- Krótkiewicz, Marek 468
- Kuczyński, Karol 173
- Kulczycki, Piotr 437
- Kulikowski, Juliusz L. 57
- Kupka, Tomasz 291
- Kuraszkiewicz, Bożena 200
- Kurzynski, Marek 421
- Kwasnicka, Halina 157
- Kwiatkowski, Jan 85
- Lamza, Aleksander 105
- Lewandowski, Zbigniew 482

- Maciejewski, Henryk 395
 Małyśiak, Bożena 377
 Mańka, Ilona 529
 Marciniak, Jan 521, 529, 537, 545
 Marnik, Joanna 505
 Mastej, Andrzej 377
 Matonia, Adam 291
 Mazurek, Urszula 387
 Mężyk, A. 551
 Michnik, R. 551
 Mikołajczak, Paweł 173, 184
 Miszczak, Jan 315
 Momot, Alina 267, 275
 Momot, Michał 275
 Mrozek, Dariusz 377
 Musioł, Joanna 387
- Nieniewski, Mariusz 129
 Nowak, Robert 357
- Ostrek, Grzegorz 139
- Pander, Tomasz 259, 283
 Paradowski, Mariusz 157
 Paszenda, Zbigniew 15, 521, 551
 Pawlus, Bartłomiej 335
 Piątek, L. 225
 Pieta, L. 66, 74
 Piętka, D. 403
 Piętka, Ewa 93, 149, 192
 Poleć, Paweł 335
 Pospiech-Kurkowska, Sylwia 453
 Prędko, Bartłomiej 497
 Przelaskowski, Artur 44, 113, 139,
 208, 325
 Przybyła, Tomasz 283
 Przytułska, Małgorzata 57
- Radomski, Dariusz 325, 482
 Rak, Z. 551
 Rekawek, Agnieszka 488
 Roj, Dawid 291
 Roszkowski, Piotr I. 482
 Rycerski, W. 551
- Sasiadek, Marek J. 157
 Sklinda, Katarzyna 139
 Słowiński, Roman 497
 Sokolowska, Beata 488
 Spinczyk, Dominik 513
 Stefko, Kamil 179
 Straszeczka, Ewa 429
 Szczepaniak, P.S. 66, 74
 Szela, Magdalena 505
 Szewczenko, Janusz 529
- Tarnawski, Wojciech 217
 Tomczyk, A. 66, 74
 Topolski, Mariusz 421
- Walczak, Wojciech 85
 Walke, Witold 521, 529, 545, 559
 Więclawek, Agnieszka 387
 Więclawek, Wojciech 93
 Wierzbicka, Diana 57
 Wojtkiewicz, Krystian 468
 Wolczowski, Andrzej 421
 Wozniak, Michał 445
 Wrobel, Janusz 291, 299, 307
 Wrobel, Zygmunt 105, 232, 568
- Zajdel, Jakub 335
 Zalewska, Ewa 315
 Zarychta, Piotr 165
 Zarychta-Bargieła, Anna 165
 Zdunczuk, Mateusz 345
 Żmudzki, Jarosław 559